

字典学习模型、算法及其应用研究进展

练秋生¹ 石保顺¹ 陈书贞¹

摘要 稀疏表示模型常利用训练样本学习完备字典,旨在获得信号的冗余稀疏表示. 设计简单、高效、通用性强的字典学习算法是目前的主要研究方向之一,也是信息领域的研究热点. 基于综合稀疏模型的字典学习方法已经广泛应用于图像分类、图像去噪、图像超分辨率和压缩成像等领域. 近些年来,解析稀疏模型、盲字典模型和信息复杂度模型等新模型的出现丰富了字典学习理论,使得更广泛类型的信号能够被“简单性”描述. 本文详细介绍了综合字典、解析字典、盲字典和基于信息复杂度字典学习的基本模型及其算法,阐述了字典学习的典型应用,指出了字典学习的进一步研究方向.

关键词 字典学习, 稀疏表示, 综合模型, 解析模型

引用格式 练秋生, 石保顺, 陈书贞. 字典学习模型、算法及其应用研究进展. 自动化学报, 2015, 41(2): 240–260

DOI 10.16383/j.aas.2015.c140252

Research Advances on Dictionary Learning Models, Algorithms and Applications

LIAN Qiu-Sheng¹ SHI Bao-Shun¹ CHEN Shu-Zhen¹

Abstract The sparse model often utilizes training samples to learn an over-complete dictionary, in order to obtain the redundant and sparse representation of signals. Designing simple, effective and flexible dictionary learning algorithms is one of the main and hot research topics in the information field. The dictionary learning methods based on synthesis sparse model have been applied into image classification, image denoising, image super-resolution and compressive imaging. In recent years, analysis sparse model, blind dictionary model and information complexity model have been proposed, which enrich the dictionary learning theory in order and lead to a “simple” description for a wide range of signals. In this paper, the fundamental models and dictionary learning algorithms are introduced in detail in terms of synthesis dictionary, analysis dictionary, blind dictionary and dictionary learning based on information complexity. Typical applications of dictionary learning methods are further illustrated. Finally, the directions for further research of the dictionary learning are pointed out.

Key words Dictionary learning, sparse representation, synthesis model, analysis model

Citation Lian Qiu-Sheng, Shi Bao-Shun, Chen Shu-Zhen. Research advances on dictionary learning models, algorithms and applications. *Acta Automatica Sinica*, 2015, 41(2): 240–260

在信号处理领域,人们一直对信号的“简单性”描述高度重视.从信息论的角度看,如果信号是稀疏的,或具有某种结构,或可用某个确定的模型来表示,这样的信号称之为“简单”信号.“简单”信号具有较少的信息量,由少量比特就可以表示.“简单性”是简单信号具有的固有特性,它通常表现为稀疏性、低秩性、低熵等.作为信号“简单性”描述的稀疏表示技术,是近些年来学者们研究的重点,发展也相对比较成熟.1959年,Hubel等^[1]研究猫的简单细胞感受野,发现位于主观视觉皮层 V1 区的细胞能够

对视觉信息进行稀疏表示,从此稀疏性研究引起了学者们^[2–3]的关注.生物学家指出哺乳类动物在长期的进化中,形成了能够快速、准确、低能耗地表示自然图像的视觉神经方面的能力,其关键在于哺乳动物感知视觉信息时,其大脑视觉皮层中只有少量神经元处于激活状态,即视觉信息可以用少量的神经元进行稀疏表示.随着稀疏表示理论研究的深入,这一理论被广泛应用于信号处理领域.从综合角度看,信号可以表示为少量字典原子的线性组合,即信号在字典表示下是稀疏的.从分析角度看,信号经过合适的变换是稀疏的,即信号在变换域是稀疏的.字典学习方法通过优化相应的字典学习代价函数,获得能够对信号进行稀疏表示的字典,或从分析的角度看,通过优化代价函数使变换系数最稀疏.

早期提出的主成分分析 (Principle component analysis, PCA)^[4]和独立成分分析 (Independent component analysis, ICA)^[5]可以看作字典学习的萌芽阶段,它们都是通过训练样本学习信号的自适

收稿日期 2014-04-14 录用日期 2014-10-12

Manuscript received April 14, 2014; accepted October 12, 2014
国家自然科学基金 (61471313), 河北省自然科学基金 (F2014203076) 资助

Supported by National Natural Science Foundation of China (61471313), and Natural Science Foundation of Hebei Province (F2014203076)

本文责任编辑 刘成林

Recommended by Associate Editor LIU Cheng-Lin

1. 燕山大学信息科学与工程学院 秦皇岛 066004

1. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004

应表示. PCA 基于最小均方误差准则和二阶统计方法, 利用训练样本协方差矩阵的特征向量来表示信号. 特征向量可看作信号的基原子, 原子之间相互正交. PCA 的主要应用包括数据压缩和模式识别系统中的特征降维. ICA 是一种高阶统计分析方法, 它可以看成是 PCA 方法的推广. ICA 基于独立性准则通过对训练样本进行学习获得一组非正交基 (原子), 它主要应用于盲源分离 (Blind source separation, BSS)^[6]. PCA 和 ICA 虽然都有字典学习的雏形, 但是两者都没有显式地利用稀疏先验. 真正意义上的字典学习方法可以追溯到上世纪 90 年代. 1993 年, Mallat 等发表了一篇奠基性论文^[7], 首次阐述了过完备字典 (原子数量比信号维数大的多) 概念, 并提出了解决过完备字典稀疏表示问题的匹配追踪算法. 随后 Chen 等于 1995 年提出了解决稀疏优化问题的基追踪方法^[8]. 这两篇论文虽然都没有涉及具体的求解字典问题的方法, 但它们提出了稀疏优化问题的有效求解算法, 从而奠定了字典学习理论的基础. 字典学习真正引起人们的广泛关注是在 1996 年 Olshausen 等发表有关自然图像稀疏编码的开创性论文^[9-11] 之后. 他们导出了 l_1 范数作为系数稀疏性度量, 令人惊奇的是这种以稀疏性为准则进行学习获得的字典其每个原子的形态与视觉皮层中 V1 区简单细胞的感受野类似. Olshausen 等的研究成果奠定了稀疏编码的神经生理学基础, 点燃了人们对字典学习的研究热情.

近些年, 以信号的稀疏性先验求解图像反问题引起了学者们的广泛关注^[12-14], 尤其是压缩感知 (Compressive sensing, CS) 领域. 根据 Donoho 和 Candes 等的理论^[12-13], 信号在字典下的表示系数越稀疏则重构质量越高, 因此字典的选择十分重要, 它决定了图像反问题的求解质量. 目前, 字典构造方法一般分为两种: 解析方法和学习方法. 基于解析方法构造的字典通过事先定义好的某种数学变换或调和分析方法来构造, 字典中的每个原子可用数学函数或少量的参数来刻画, 如离散余弦变换、小波变换、双树复数小波变换、轮廓波变换、Shearlet、Grouplet^[15] 以及参数化字典^[16] 等. 该方法虽然构造相对简单, 计算复杂度低, 但原子的基本形状固定, 原子的形态不够丰富, 不能与图像本身的复杂结构最佳匹配, 即非最优表示. 近年来人们开始根据数据或信号本身来学习过完备字典, 这类字典中的原子与训练集中的信号本身相适应. 与基于解析方法的字典相比, 通过学习获得的字典原子数量更多, 形态更丰富, 能更好地与信号或图像本身的结构匹配, 具有更稀疏的表示. 近几年的研究成果表明, 通过学习获得的字典比解析方法构造的字典在图像去噪、图像修复、图像超分辨率等方面有更

出色的性能^[17-19].

近些年, 以设计通用性强、简单、高效的字典学习算法为目标^[20] 的字典学习方法迅猛发展, 基于综合模型的约束字典、在线字典、结构字典等^[21-23] 在数据或字典结构上添加约束提升了字典学习的效率. 字典学习模型也不再局限于综合稀疏模型, 解析模型、盲字典模型、信息复杂度模型等在分析问题的角度、“简单性”度量等方面拓展了综合稀疏模型, 使得字典学习模型更加多元化. 基于新型模型的字典学习方法丰富了字典学习理论, 扩充了应用范围.

本文在第 1 节介绍了字典学习的基本原理; 第 2 节对综合字典学习算法的基本模型及关键技术进行了详细地介绍; 第 3 节阐述了解析字典学习方法的相关模型及其研究成果; 第 4 节研究了基于盲字典模型的盲字典学习方法; 第 5 节讨论了基于信息复杂度模型的字典学习方法; 第 6 节概述了字典学习方法的应用; 最后对全文进行了总结并对字典学习的研究前景进行了展望.

1 字典学习的基本原理

信号的“简单性”描述是学者们研究的重点, 尤其是信号的稀疏性描述, 即稀疏表示. 目前, 在信息领域, 稀疏表示的研究被分为两种模型^[24]: 综合模型 (Synthesis model) 和解析模型 (Analysis model). 综合模型给定字典 $D = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K] \in \mathbf{R}^{n \times K}$, 字典 D 中的每一列 $\mathbf{d}_k \in \mathbf{R}^n$ 称为一个原子, 则信号 $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbf{R}^{n \times N}$ 可以表示为字典 D 中若干原子的线性组合: $X \approx DA$, 其中 $A = [\alpha_1, \alpha_2, \dots, \alpha_N] \in \mathbf{R}^{K \times N}$ 为信号 X 在字典 D 下的表示系数. 如果 α_i 中仅有 s ($s \ll K$) 个非零系数, 则称系数 α_i 是 s -稀疏的. 综合模型一般情况下使用 l_0 或 l_1 范数衡量表示系数的稀疏性, 由于非参数贝叶斯字典学习模型本质上属于综合模型, 但采用与范数稀疏性度量类似的贝塔-贝努利过程度量, 故本文用 $g_A(\cdot)$ 统一度量综合模型中系数的稀疏性, 所对应的字典学习优化问题可描述为

$$\min_{D, A} \{ \|X - DA\|_F^2 + \lambda_A g_A(A) \} \quad (1)$$

式中, X 表示训练集, D 为学习字典, A 为所对应的系数矩阵, λ_A 为正则化参数. 与综合模型所对应的是解析模型, 基于解析模型的稀疏表示研究还处于起步阶段, 该模型也称为共稀疏解析模型 (Cosparse analysis model)^[25-26], 在共稀疏解析模型中, 假设训练集 X 是共稀疏的, 学习解析字典 $\Omega = [\omega_1, \omega_2, \dots, \omega_K]^T \in \mathbf{R}^{K \times n}$ ($K \geq n$) 的优化问题可描述成:

$$\hat{\Omega} = \arg \min_{\Omega} g(\Omega X) \quad (2)$$

其中, g 是度量 ΩX 稀疏度的函数, 即:

$$g(\Omega X) = \|\Omega X\|_p \quad (3)$$

式 (3) 中, p 范数可以是 l_0 、 l_1 范数等. 该模型衡量 ΩX 的稀疏性采用的是共稀疏度 u , 即零元素的个数: $u = K - \|\Omega X\|_0$, 共稀疏度越大, ΩX 越稀疏. 图 1 为基于过完备字典的两种模型原理示意图, 图 1 中与非零系数对应的原子用虚线框标出. 从图 1(a) 的综合模型示意图中可以看出信号 x 在字典 D 下是 3 稀疏的, 即信号 x 可由字典 D 的三个原子进行逼近, 从图 1(b) 的解析模型示意图中可以看出, ΩX 有 3 个非零系数, 9 个零系数, 共稀疏度为 9.

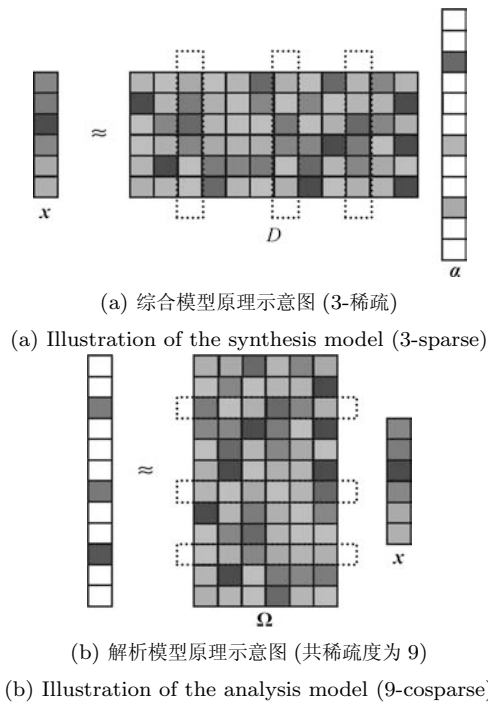


图 1 字典稀疏表示原理

Fig. 1 Sparse representation based on dictionary

上述字典学习模型都有一个前提条件: 样本本身的信息是完整或基本完整的. 样本信息基本完整是指样本受到强度较小的噪声污染 (包括高斯噪声和脉冲噪声等) 或少量的像素缺失. 最近 Gleichman 等^[27] 提出的盲压缩感知 (Blind compressed sensing, BCS) 理论, 也即盲字典学习, 打破了这一限制, 为稀疏表示领域的研究者提供了一个全新的视野. BCS 要解决的问题是: 假设信号矩阵 X 中的每一列 x_i 能用某个字典 D 进行 s -稀疏表示, 给定测量矩阵 $\Phi \in \mathbf{R}^{m \times n}$ ($m \ll n$) 及测量值矩阵 $Y = \Phi X = [y_1, y_2, \dots, y_i, \dots]$, 在字典 D 未知的情况下重构 X , 盲字典学习模型为

$$\min_{D, A} \{\|Y - \Phi D A\|_F^2 + \lambda g_A(A)\} \quad (4)$$

其中, 第一项是信号的保真项, 第二项是信号的稀疏正则项, 正则化参数 λ 用以控制两项的比例.

如果信号是“简单”的、复杂度低、包含的信息量少, 那么用少量的比特就能表示它. 基于这种认识, Jalali 等建议用信息论中 Kolmogorov 复杂度来统一表示信号的稀疏性、结构性和低秩性^[28], 但是 Kolmogorov 复杂度是不可计算的. 一种可行的方法是用最小描述长度 (Minimum description length, MDL) 代替 Kolmogorov 复杂度. 可压缩性能够很好地间接衡量模型刻画信号规律的能力^[29], MDL 框架基于这一认识训练字典. 文献 [29] 以最小码长作为衡量模型的准则, 在模型集合中寻找最佳模型. 他们将综合模型中满足稀疏条件的一对 $\{A, D\}$ 看作是模型集合 $M(s, K) = \{(A, D) : \|\alpha_i\|_0 \leq s, i = 1, \dots, N, D \in \mathbf{R}^{n \times K}\}$ 中的一个元素 $M = (A, D)$, 其中 s 表示稀疏度, 将基于最小描述长度准则的字典学习问题看做模型的选择问题: 给定训练集 X , 从模型集合 $M(s, K)$ 中选择最佳模型 $\hat{M} = (\hat{A}, \hat{D})$. 用码长 L 的最小化作为最小描述准则:

$$\min_{A, D} \{L(X, A, D) : (A, D) \in M(s, K)\} \quad (5)$$

MDL 框架在字典学习过程中同时优化了稀疏度和字典规模:

$$\min_{s, K} \{L(X, \hat{A}(s, K), \hat{D}(s, K)) : 0 \leq s \leq K, K > 0\} \quad (6)$$

MDL 框架在字典学习过程中, 求解了优化问题 (6), 能够自动地更新稀疏度、字典规模等参数, 具有独特的优势.

本文将字典学习模型分成了 4 种模型进行介绍, 这 4 种模型既有区别又有联系, 本质上它们都可以统一到稀疏表示的框架中. 综合字典学习模型试图找到一组能够反映信号本征空间的基, 而解析字典学习模型则是从对偶分析的角度考虑稀疏表示问题, 它试图找到信号正交空间的基. 这两种模型的训练样本为特定的信号库, 而盲字典学习模型的训练样本是待重构信号的测量值 (待重构信号的降维投影). 三种模型均以 l_p 范数描述信号的稀疏性, 而基于信息复杂度的字典学习模型则利用信息的复杂度描述信号的“简单性”. 信号越稀疏, 则其复杂度越低, 因此它仍然属于广义的稀疏表示模型.

2 综合字典学习算法

综合稀疏模型一直是学者们关注的重点, 基于这一模型的字典学习方法也是字典学习理论研究中相对比较成熟的部分, 大部分算法求解式 (1) 都采

用系数更新和字典更新交替优化的方式. 这些算法的区别主要在于字典更新方式, 而系数更新部分则大多数算法没有本质区别, 固定字典 D 更新稀疏系数 A 是标准的稀疏编码问题. 从理论上说, 任何一种稀疏编码方法都可以用于系数更新. 近二十年来, 人们提出了大量的稀疏编码方法^[30-36]. 固定稀疏系数 A 更新字典 D 是研究字典学习算法的学者们最为关注的环节.

2.1 Sparsenet 字典学习算法

1996 年, Olshausen 等在《Nature》上发表了著名的 Sparsenet 字典学习算法^[9], 该算法奠定了字典学习理论的基础. 在文献 [10-11] 中他们对 Sparsenet 算法进行了详细的论述和扩充. Sparsenet 算法利用最大似然估计学习字典, 该算法从自然图像库中抽取大量的小图像块作为训练集 X , 每一个小图像块 \mathbf{x} 满足如下模型:

$$\mathbf{x} = D\boldsymbol{\alpha} + \mathbf{v} \quad (7)$$

式中, \mathbf{v} 表示残差向量, 其每一个分量服从方差为 σ^2 的高斯分布. 利用最大似然方法求解最优字典 D , 即寻找字典 D 使得似然函数 $P(X|D)$ 最大. 假设训练集 X 中每一小图像块 \mathbf{x}_i 都是统计独立的, 即: $P(X|D) = \prod_{i=1}^N P(\mathbf{x}_i|D)$, 则字典学习对应的优化问题可表示为

$$\max_D P(X|D) = \max_D \prod_{i=1}^N P(\mathbf{x}_i|D) \quad (8)$$

由统计学知识, 上式等价于

$$\max_{D, \boldsymbol{\alpha}_i} \left\{ \prod_{i=1}^N P(\mathbf{x}_i|\boldsymbol{\alpha}_i, D) P(\boldsymbol{\alpha}_i) \right\} \quad (9)$$

上式中的 $P(\mathbf{x}_i|\boldsymbol{\alpha}_i, D)$ 可由式 (7) 得: $P(\mathbf{x}_i|\boldsymbol{\alpha}_i, D) \propto e^{-\|D\boldsymbol{\alpha}_i - \mathbf{x}_i\|_2^2 / 2\sigma^2}$, $P(\boldsymbol{\alpha}_i)$ 表示系数的先验分布, 文献 [11] 中假设 $P(\boldsymbol{\alpha}_i) \propto \gamma_C(1 + \boldsymbol{\alpha}_i^2)^{-1}$, 则式 (9) 变为

$$\max_{D, \boldsymbol{\alpha}_i} \left\{ \prod_{i=1}^N e^{-\frac{\|D\boldsymbol{\alpha}_i - \mathbf{x}_i\|_2^2}{2\sigma^2}} \gamma_C(1 + \boldsymbol{\alpha}_i^2)^{-1} \right\} \quad (10)$$

对式 (10) 取自然对数后将负号去掉, 从而使最大化变为最小化, 再将系数合并后即可获得最终的优化问题:

$$\min_{D, \boldsymbol{\alpha}_i} \left\{ \sum_{i=1}^N \|D\boldsymbol{\alpha}_i - \mathbf{x}_i\|_2^2 + \lambda_A \ln(1 + \boldsymbol{\alpha}_i^2) \right\} \quad (11)$$

利用交替优化方法求解上述问题, 固定字典 $D^{(t-1)}$, 其中 t 表示迭代次数, 用简单的梯度下降法

求解表示系数, 固定系数 $\boldsymbol{\alpha}_i^{(t)}$ 利用下式更新字典:

$$D^{(t)} = D^{(t-1)} - \eta \sum_{i=1}^N \{(D^{(t-1)}\boldsymbol{\alpha}_i^{(t)} - \mathbf{x}_i)(\boldsymbol{\alpha}_i^{(t)})^T\} \quad (12)$$

其中, η 表示步长. 除了上述推导过程给出的系数先验分布满足柯西分布外, 系数先验分布还可以假设为拉普拉斯分布: $P(\boldsymbol{\alpha}_i) \propto e^{-\gamma_L|\boldsymbol{\alpha}_i|}$, 导出的优化问题以 l_1 范数度量系数的稀疏性^[10, 37]. Olshausen 等^[10] 发现以稀疏性为准则学习的字典原子具有局部化、方向性和带通性的特点, 并且一些栅格状的原子能够有效地匹配图像的纹理特征. 受 Olshausen 等工作的启发, Kreutz-Delgado 等提出了最大后验概率字典学习算法^[38], 其优化的代价函数是使字典 D 出现的后验概率 $P(D|X)$ 最大.

上述算法容易陷入局部最优, 为训练全局最优字典, Mailhé 等对 Olshausen 的算法进行了改进^[39], 他们将梯度下降中的固定步长改为自适应步长, 从而越过局部最优点, 最终以较大的概率保证在全局最优点收敛. 为提高 Olshausen 字典学习算法的速度, Lee 等^[40] 将字典学习代价函数的优化转化为交替求解 l_1 范数正则化最小均方问题和 l_2 范数约束的最小均方问题, 并提出了特征符号搜索算法和拉格朗日对偶算法分别求解这两个凸优化问题. Lee 的算法适合从大规模训练集中学习高冗余的过完备字典. 令人感兴趣的是他们发现用这种方法进行训练获得的字典原子除了呈现 V1 区简单细胞感受野的特性外, 还有部分原子的特性与 V1 区的 End-stopping 细胞的感受野类似.

2.2 MOD 与 K-SVD 字典学习算法

早期的 Sparsenet 字典学习算法及其改进算法, 大都受视觉系统的工作机制启发, 并且多数利用梯度下降算法求解所对应的优化问题. 受广义 Lloyd 聚类算法的启发, Engan 等对 Olshausen 的 Sparsenet 字典学习算法进行了改进, 提出了 MOD (Method of optimal directions) 字典学习算法^[41]. MOD 算法采用 l_0 范数衡量信号的稀疏性, 利用交替优化方式求解式 (1). 但它在更新字典时定义整体表示均方误差:

$$\|E\|_F^2 = \|[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N]\|_F^2 = \|X - DA\|_F^2 \quad (13)$$

采用最小均方误差算法更新字典:

$$D^{(t)} = X(A^{(t-1)})^T (A^{(t-1)}(A^{(t-1)})^T)^{-1} \quad (14)$$

该算法的缺点是需要大矩阵的乘法和求逆计算, 对存储容量的要求高. 为减小 MOD 算法的复

杂度, Aharon 等提出了 K-SVD (K-singular value decomposition) 算法^[20], 所对应的优化问题为

$$\min_{D,A} \{\|X - DA\|_F^2\} \quad \text{s.t.} \quad \|\alpha_i\|_0 \leq s, \forall i \quad (15)$$

固定字典 $D^{(t-1)}$, 采用正交匹配追踪 (Orthogonal matching pursuit, OMP)^[32] 算法求解以下优化问题:

$$\min_{\alpha_i} \{\|\mathbf{x}_i - D^{(t-1)}\alpha_i\|_2^2\} \quad \text{s.t.} \quad \|\alpha_i\|_0 \leq s, \forall i \quad (16)$$

该算法与 MOD 算法不同之处在于字典更新方式, K-SVD 算法更新字典时不是对整个字典一次更新, 而是逐个原子更新. 定义移除第 k 个原子后信号表示的误差矩阵 E_k :

$$E_k = X - \sum_{j \neq k} d_j \alpha_T^j \quad (17)$$

式中, α_T^j 为 A 的第 j 个行向量. 对 E_k^R 进行奇异值分解, 即 $E_k^R = U\Sigma V^T$, 更新字典原子 $\hat{d}_k = u_1$, 同时更新系数为: $\hat{\alpha}_k^R = \Sigma[1, 1]v_1$. K-SVD 算法不需要矩阵求逆计算, 运算复杂度比 MOD 算法低得多, 并且在字典更新步骤中对系数矩阵与字典原子联合更新, 提高了算法收敛速度, 它在实际中应用最广泛. 随后, 许多学者对其进行了改进, Smith 等^[42] 在字典学习代价函数 (15) 中加入支撑完整的先验信息, 提出了多重字典更新 (Multiple dictionary update, MDU) 算法, 该算法利用上次迭代的稀疏表示残差信息进行系数更新, 能够提升字典训练速度与性能. Rubinstein 等^[43] 在字典更新步骤中, 不再对每个字典原子用 SVD 方法寻找的秩 -1 逼近, 而是用交替优化方法求解以下联合优化问题:

$$\min_{d_k, \alpha_T^k} \{\|E_k^R - d_k \alpha_T^k\|_F^2\} \quad \text{s.t.} \quad \|d_k\|_2 = 1 \quad (18)$$

该方法避免了大量的矩阵分解, 比原始 K-SVD 方法字典训练效率更高, 但图像去噪效果却有所下降. Sadeghi 等^[44] 提出基于并行原子更新的字典学习算法, 该算法对 K 个字典原子同时进行更新, 在每次迭代过程中, 更新字典后的误差用以下秩 -1 形式计算: $E = E_k - d_j \alpha_T^j$, 该计算方式使得更新原子能够并行处理. 基于并行字典原子更新的方法不仅提升了速度, 还提高了性能. Dai 等推广了 K-SVD 的思想, 提出了同步码字 (原子) 优化 (Simultaneous codeword optimization, SimCO)^[45] 算法. 该算法能够同时更新字典中多个原子和对应的稀疏系数, 在学习速度和字典性能两方面都优于 K-SVD 算法. 上述几种字典学习算法中都需要预先确定字典规模 (即字典中原子的数量), 但在实际应用中字典规模的最优值的选取没有相应的理论指导. 为解决这个问

题, 分段 K-SVD 算法^[46] 在学习过程中逐步增大字典规模, 直到表示误差达到要求为止, 达到了自适应选择字典规模的目的; 尺度自适应字典学习算法^[47] 将多变量莫罗邻近示性函数 (Multivariate Moreau proximal indicator, MMPI) 作为惩罚项加入到字典学习的代价函数中, 优化了字典规模, 对于训练大字典效果更好、速度更快. MOD、K-SVD 算法及其改进算法的思想均源于 K 均值 (K -means) 聚类方法, 但 K-SVD 算法在字典更新步骤中会破坏稀疏系数的结构, MOD 算法虽然没有破坏稀疏系数的结构, 但原子更新不是顺序更新的. 针对这些问题, Sahoo 等^[48] 提出了 SGK (Sequential generalization of K -means) 算法, 该算法具有不破坏稀疏系数的结构的优点, 不仅能够逐原子更新, 而且算法复杂度比 K-SVD 低. Sadeghi 等^[49] 根据递归迭代的思想, 将式 (15) 中的字典和系数乘积进行一阶展开, 得到了新的字典学习代价函数:

$$\min_{D,A} \{\|X - D^{(t-1)}A^{(t-1)} - DA^{(t-1)} - D^{(t-1)}A\|_F^2\} \quad \text{s.t.} \quad \|\alpha_i\|_0 \leq s, 1 \leq i \leq N \quad (19)$$

文献 [49] 的实验结果表明, 利用现有的 MOD, MDU, SGK 算法去求解式 (19) 比式 (15) 能够更准确、更快地训练字典.

2.3 具有多种约束条件的字典学习算法

在字典学习中, 人们期望字典和字典原子具有一些特定的性质或约束条件, 例如字典原子单位范数约束、字典 Frobenius 范数有界性约束、相关性约束等. 施加约束条件的字典学习优化问题相当于在式 (1) 的基础上加入与字典约束条件对应的正则项 $g_D(\cdot)$:

$$\min_{D,A} \{\|X - DA\|_F^2 + \lambda_A g_A(A) + \lambda_D g_D(D)\} \quad (20)$$

其中, λ_D 为正则化参数. Yaghoobi 利用 MM (Majorization minimization) 算法^[21] 处理以上优化问题. 以字典 Frobenius 范数有界性为例, 将这种限制以朗格朗日形式加入上述优化问题:

$$\min_{D,A} \{\|X - DA\|_F^2 + \lambda_A g_A(A) + \lambda_D (\|D\|_F^2 - c_F)\} \quad (21)$$

式中, c_F 为常数. 在更新字典时 MM 算法用一个代理函数代替问题 (21) 的代价函数:

$$\min_D \{\|X - DA^{(t-1)}\|_F^2 + \lambda_D (\|D\|_F^2 - c_F) + C_D \|D - D^{(t-1)}\|_F^2 - \|DA^{(t-1)} - D^{(t-1)}A^{(t-1)}\|_F^2\} \quad (22)$$

其中, $C_D > \|(A^{(t)})^T A^{(t)}\|_2$ 为一个常数, 该算法利用邻近原子来求解问题 (22), 比 K-SVD 算法的收敛速度更快. Rakotomamonjy 提出基于交替方向乘子法 (Alternating direction method of multipliers, ADMM) 的字典学习算法^[50], 该算法解决了式 (20) 中加入字典原子的单位范数、非负性等多个约束条件的优化问题. 文献 [50] 给出的实验结果表明在训练样本有噪声干扰的情况下, 同时考虑多个原子约束条件的 ADMM 算法学习的字典其信号表示误差比 MM 算法和 K-SVD 算法更小. 在此基础上, 文献 [51] 提出了解决多重约束字典学习问题的直接优化算法, 该算法与常规算法的最大区别是迭代时不再交替优化字典和表示系数, 而是用邻近梯度下降法直接优化式 (20), 同时优化字典和表示系数.

相干性约束是过完备字典学习算法中需要重点考虑的一个因素. 正交字典的相干性为 0, 而过完备字典由于原子数量大于信号维数, 原子之间的相干性大于 0. 一般来说, 字典的冗余度越大其相干性也越大, 字典的相干性对其稀疏编码性能十分关键. 在相同冗余度的前提下, 低相干性可加快稀疏编码速度, 改善信号重构性能. 另外低相干性还有助于避免字典原子对训练样本的过拟合, 避免字典中出现过于相似的原子对. 利用等角度紧标架 (Equiangular tight frame, ETF) 具有最小相干性的特点^[52], Sigg 等在字典更新时将 ETF 约束加入到代价函数:

$$\min_D \{ \|X - DA\|_F^2 + \gamma_D \|D^T D - I\|_F^2 \} \quad (23)$$

式中, γ_D 为正则化参数. 利用大规模有限存储拟牛顿法 (Limited-memory BFGS, LM-BFGS) 近似求解上述优化问题. 该字典学习方法能够有效控制训练字典的相干性. 而文献 [53–54] 设计低相干性字典不同于上述方法, 在训练字典时, 其对应的优化问题为

$$\begin{aligned} \min_{D \in \Upsilon} & \|D - \bar{D}\|_F^2 \\ \text{s. t. } \Upsilon = & \{D | \mu(D) \leq \bar{\mu} \wedge \|d_k\|_2^2 = 1, 1 \leq k \leq K\} \end{aligned} \quad (24)$$

其中, $\bar{\mu}$ 为指定的相关性门限, \bar{D} 为相干性小于 $\bar{\mu}$ 的参考字典. 求解上述问题时, 通常在字典更新步骤中, 对原子进行操作, 使得原子满足上述要求. Mailhe 等提出的非相干 K-SVD (Incoherent K-SVD, INK-SVD) 字典学习算法^[53], 在更新字典时, 如果发现某一对原子的相干性大于 $\bar{\mu}$, 则对这两个原子进行旋转以降低它们的相干性. 这种降低相关性操作迭代进行, 直到字典中任何两个原子的相干性均不大于指定的门限. Barchiesi 等提出的迭代投影和旋转的低相关性字典学习算法^[54] 与 INK-

SVD 算法的思想类似, 与 INK-SVD 不同的是在更新字典时, 它们对整个字典而不是原子对进行降低相关性和旋转操作.

2.4 在线字典学习算法

在 Sparsenet、MOD、K-SVD 等字典学习算法中, 每次迭代更新字典时需要用所有的训练样本进行计算, 对存储空间及计算能力要求高, 不适合大规模样本和动态到达样本的字典学习. 针对这个问题, Mairal 等提出了每次迭代只处理一个样本的在线字典学习 (Online dictionary learning, ODL)^[22] 算法, 该算法每次从样本集中随机取出一个样本 x_i 进行字典训练, 根据上次迭代的字典 $D^{(t-1)}$ 用 LARS (Least-angle regression) 算法进行稀疏编码, 通过求解下式获得表示系数 α_t :

$$\min_{\alpha} \{ \|x_i - D^{(t-1)} \alpha\|_2^2 + \lambda_{\alpha} \|\alpha\|_1 \} \quad (25)$$

固定系数 α_t , 更新字典:

$$\min_{D \in \zeta} \left\{ \frac{1}{t} (\text{tr}(D^T D B_t) - \text{tr}(D^T C_t)) \right\} \quad (26)$$

其中, $\zeta := \{D \in \mathbf{R}^{n \times K}, \text{ s. t. } \forall k = 1, \dots, K, d_k^T d_k \leq 1\}$, $B_t = B_{t-1} + \alpha_t \alpha_t^T$, $C_t = C_{t-1} + x_t \alpha_t^T$, 利用块坐标下降算法求解式 (26) 来更新当前字典. Mairal 等在文中证明了 ODL 算法的收敛性, 给出的实验结果表明, ODL 算法比批量处理的字典学习算法收敛速度更快, 能够实现百万级大规模样本字典学习. 在文献 [55] 中, Skretting 等通过改进 MOD 算法中的字典更新方式, 提出了 RLS-DLA (Recursive least squares-dictionary learning algorithm). RLS-DLA 采用一个样本更新一次字典的策略, 并利用 Woodbury 矩阵恒等式推导出字典更新递推公式, 避免了 MOD 算法中的矩阵求逆计算. 该算法比 ODL 的计算复杂度更低, 收敛速度更快. 在文献 [56] 中 Labusch 等提出了追踪包 (Bag of pursuits, BOP) 稀疏编码算法, 利用 BOP 和随机梯度下降法提出了适合在线学习的字典学习方法. 该算法简单、快速、鲁棒, 即使在样本较少的情况下仍能训练出正确的字典.

2.5 结构字典学习算法

人们发现利用小波的结构特性处理图像反问题的效果往往要比纯粹用稀疏性的效果有明显提升^[57]. 受正交小波字典的启发, 近几年人们开始关注结构字典学习. 目前结构字典主要包括: 块稀疏字典^[23]、正交基联合字典^[58]、广义主成分分析^[59]、组结构字典^[60–61]、树结构字典^[62]、多尺度字典^[63–64]、多级字典^[65]、双稀疏字典^[66]、可压缩字典^[67]、可分离字典^[68] 和平移不变字典^[69–72] 等. 块稀疏字典是

指字典具有某种块结构, Zelnik-Manor 等^[23] 假设信号可以用少数几个子空间联合表示, 提出了块稀疏字典学习算法. 该算法试图找到能够对信号 X 进行 q -块稀疏表示的字典 \hat{D} 和块结构 \hat{G} :

$$\{\hat{D}, \hat{G}, \hat{A}\} = \arg \min_{D, G, A} \|X - DA\|_F$$

$$\text{s. t. } \|\alpha_i\|_{0, G} \leq q, 1 \leq i \leq N, |G_p| \leq s_{\max}, p \in G \quad (27)$$

其中, s_{\max} 为最大块的大小, $|G_p|$ 表示第 p 块中包含的原子个数, 即块的大小. 所谓 q -块稀疏信号意味着信号的非 0 表示系数集中在 q 块中出现, 每块包括多个系数, 每块的尺寸可以相同也可以不同, 但块与块之间没有重叠, 块结构 G 需要在学习过程中训练. 图 2 给出了一个块稀疏向量与所对应的块结构字典的例子. 向量 α_i 分成了 4 个块, 其中阴影部分表示向量中的非零元素, 总共 8 个非零元素, 占了两个块, 为 2-块稀疏信号, 即 $q = 2$.

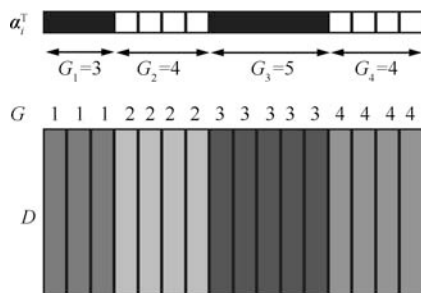


图 2 块稀疏向量与所对应的块结构字典

Fig. 2 Block sparse vector and its corresponding block structured dictionary

Zelnik-Manor 等^[23] 采用交替优化方式求解式 (27). 首先固定字典 D , 用 SAC (Sparse agglomerative clustering) 算法根据字典原子所表示样本的相似性逐步对原子进行合并, 从而更新块结构 G 和表示系数 A ; 其次, 固定块结构 G , 用 K-SVD 的一种推广形式 BK-SVD (Block K-SVD) 算法更新字典 D 和表示系数 A . 实验结果表明, 对于块稀疏数据, 块结构字典比没有利用块结构学习的字典对相同数据的表示误差更低, 重建效果更好. 正交基联合字典^[58] 和广义主成分分析^[59] 可以看作块稀疏字典的两个特例. 正交基联合字典由多个正交基联合组成: $D = [D_1, D_2, \dots, D_j, \dots]$, 而基于广义主成分分析的字典学习算法假设信号是由多个低维子空间中的其中一个生成的, 即信号的块稀疏度为 1. 组结构字典与块稀疏字典有类似之处, 但组结构比块结构更为通用也更灵活. 组稀疏不仅包括块稀疏, 也包括树结构稀疏, 另外组与组之间允许部分重叠. Bengio 等^[60] 以 l_1/l_2 范数做为组稀疏度量 (组内采

用 l_2 范数度量, 组与组之间采用 l_1 范数度量), 训练用于图像分类的组结构字典, 获得了良好的分类效果. Szabo 等提出了在线组结构字典学习算法^[61], 它允许组与组之间有重叠. 该算法将非凸的字典学习代价函数优化问题分解为多个凸优化问题, 用块坐标下降法和变分法推导出字典在线更新规则. 树结构字典是指字典中的原子按树状结构排列, 每个节点代表一个原子. 在稀疏编码时, 如果选择了某个原子表示信号, 则意味着该节点的所有祖先节点都要选中; 反之, 如果某个原子没有被选中, 则其所有的后代节点也不被选中. Jenatton 等将树结构稀疏性度量代替字典学习代价函数中的稀疏正则项^[62], 用原始对偶法 (Primal-dual) 求解树结构稀疏正则项对应的邻近算子, 通过交替优化方式训练出树结构字典. 多尺度字典包括 Mairal 等提出的基于空域和 Ophir 等提出的基于小波域的多尺度字典^[63-64]. Ophir 等先对训练集中的每幅图像进行小波变换, 对所有图像的同一子带抽取大量的小块用 K-SVD 算法进行训练获得该子带字典. 根据小波变换的自然属性, 不同子带的字典尺度不同. 多级字典^[65] 由多级子字典构成, 用这些子字典对信号进行分级稀疏表示, 每级子字典只选一个原子表示信号. 对样本进行稀疏编码时, 先从第一级子字典中用类似于匹配追踪的策略选出一个原子对样本进行表示, 再从第二级字典选一个原子对残差进行表示, 依此类推直到残差的能量小于指定的门限. 文献 [65] 给出实验结果说明多级字典在图像压缩及压缩感知方面的性能明显高于 K-SVD 字典.

除了信号在字典下稀疏表示的假设, Rubinstein 等提出的双稀疏字典和 Yaghoobi 等提出的可压缩字典还附加了字典原子可用某个已知的基字典或母字典中的原子进行稀疏表示的约束条件^[66-67], 即 $D = \Psi\Theta$, 其中 Ψ 基字典, Θ 为原子稀疏表示矩阵. 双稀疏字典用 l_0 范数度量 Θ 的稀疏性, 而压缩字典则用 l_1 范数度量 Θ 的稀疏性. 这两种字典学习算法学习的目的是根据双稀疏性约束条件从训练样本中得到稀疏表示矩阵 Θ . 由于可以选用存在快速算法的离散余弦基或正交小波基作为基字典, 双稀疏和可压缩字典的最大优势是稀疏编码的速度快. 可分离字典^[68] 是指字典 D 可以分离成两个较小字典的克罗内克积 $D = P_s \otimes Q_s$, 所谓的小字典指的是相对于原始字典 D 行数和列数较小, Simon 等结合字典的矩阵流形结构和几何共轭梯度法, 提出了能够对较大维数信号进行训练的字典学习算法, 该算法能够高效地进行信号重建, 而且可控制字典 D 的相干性.

平移不变字典学习算法通常是先训练出一个小的字典 D_s , 而最终的字典 D 由 D_s 中的原子通

过所有可能的平移操作获得, 因此字典 D 的冗余度非常高, 如 Jost 等用 MoTIF (Matching of time invariant filers) 算法获得的空域平移变换字典^[69]; Aharon 等提出特征字典学习算法, 先从训练集中学习出一个大尺寸的祖先原子 (Ancestral atom)^[70], 再用小尺寸的滑动窗口从祖先原子抽取大量的小图像块 (原子) 构成最终的平移不变字典. 托普利兹结构字典是一种特殊的循环字典结构, 文献 [71] 利用循环矩阵特征值分解的性质, 结合最小二乘法和矩阵向量快速算法, 提出了基于循环和托普利兹结构字典的高效字典学习方法. 相位不变字典作为平移不变字典的一种特殊结构, 是由字典生成原子通过循环移位得到的字典, 其字典原子与生成原子的相位相同. 文献 [72] 结合 K-SVD 字典学习原理, 通过最小二乘法解决了基于相位不变结构的字典更新优化问题, 由于该字典只保存生成原子并且矩阵相乘计算采用傅里叶变换快速算法, 降低了存储量和计算时间, 节省了硬件及计算开销.

2.6 非参数贝叶斯字典学习算法

大部分字典学习算法需要预先设定或利用交叉验证方式确定字典规模、稀疏度、正则化参数等信息. 这些信息对字典学习的质量非常关键, 但目前并没有完善的理论来指导, 在实际应用中大都凭经验设置和调整. 为克服这些局限, Zhou 等提出了基于截尾贝塔-贝努利过程的非参数贝叶斯字典学习算法^[73], 该算法采用如下模型:

$$\mathbf{x} = D\mathbf{h} + \mathbf{v} \quad \text{s.t.} \quad \mathbf{h} = \mathbf{w} \odot \boldsymbol{\alpha} \quad (28)$$

式中, \odot 表示哈达玛向量积, $\mathbf{w} \in \{0, 1\}^K$ 中的元素对应字典原子是否被使用, 在字典学习过程中, 该模型假定表示系数 $\boldsymbol{\alpha}$ 、字典中各个原子 \mathbf{d}_k 以及噪声 \mathbf{v} 均服从特定概率分布. 该模型通过 \mathbf{w} 衡量表示系数的稀疏性, 即:

$$\begin{aligned} \mathbf{h} = \mathbf{w} \odot \boldsymbol{\alpha} \quad \text{s.t.} \quad \mathbf{w} &\sim \prod_{k=1}^K \text{Bernoulli}(\pi_k) \\ \pi &\sim \prod_{k=1}^K \text{Beta}\left(\frac{a}{K}, \frac{b(K-1)}{K}\right) \end{aligned} \quad (29)$$

其中, π_k 为 $\boldsymbol{\pi}$ 的第 k 个分量, a, b 为模型参数. 与其他模型不同的是该算法将稀疏性先验用以上贝塔-贝努利先验假设代替, 该先验假设使得 $\boldsymbol{\alpha}$ 的许多分量恰好为 0, 本质上, 类似于 l_0 范数正则化. 文献 [73] 通过以上模型及假设, 结合贝叶斯学习自动推断出正则化参数、字典规模和重构残差, 完美地解决了字典学习中的参数选择问题.

3 解析字典学习算法

解析模型假定信号 X 是由它在某个矩阵 Ω 下的变换系数 ΩX 决定的. 与综合模型相比, 在相同维数情况下, 解析模型的子空间数量更多, 有更丰富和更灵活的表示能力^[24]. 基于这个原因, 一些学者认为解析模型在处理图像反问题时有更大的优势, 这也是近几年解析字典学习引起一些学者关注的原因^[74-76]. 不同于 Cospase 解析模型假定信号在变换域是严格稀疏的思想, Ravishankar 等首次在文献 [74] 中系统阐述了稀疏变换学习模型的概念, 他们提出的稀疏变换模型假设信号在变换域是近似稀疏的, 能够对更广泛的信号进行建模, 因此他们将稀疏变换模型看作是广义的解析模型.

3.1 Analysis K-SVD 算法

Rubinstein 等^[75] 根据信号表示的解析模型, 提出了 Analysis K-SVD 算法. Analysis K-SVD 算法对加噪后的信号进行训练, 自适应地学习解析字典并恢复原始信号, 加噪信号模型为

$$\mathbf{z}_i = \mathbf{x}_i + \mathbf{v} \quad \text{s.t.} \quad 1 \leq i \leq N \quad (30)$$

其中, \mathbf{v} 表示高斯白噪声. 该算法假设信号 \mathbf{x}_i 存在于 r ($r \ll n$) 维子空间中并且是共稀疏的, 附加解析原子单位范数约束条件, 通过求解以下优化问题, 获得解析字典 Ω :

$$\begin{aligned} \{\hat{\Omega}, \hat{X}, \{\hat{\Lambda}_i\}_{i=1}^N\} &= \arg \min_{\Omega, X, \{\Lambda_i\}_{i=1}^N} \{\|X - Z\|_F^2\} \\ \text{s.t.} \quad \Omega_{\Lambda_i} \mathbf{x}_i &= 0, \quad \text{rank}(\Omega_{\Lambda_i}) = n - r, \\ \|\omega_k\|_2 &= 1, \quad 1 \leq k \leq K \end{aligned} \quad (31)$$

式中, $Z = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N] \in \mathbf{R}^{n \times N}$, ω_k 为第 k 个解析字典原子, Λ_i 表示共支撑集, 解析模型通过共支撑刻画信号. 共支撑集决定了信号的正交空间, 即 Ω 中有 $|\Lambda_i|$ 个行与信号 \mathbf{x}_i 正交, Ω_{Λ_i} 是由 Ω 中行组成的子矩阵, 这些行在 Ω 中的索引值集合为 Λ_i . Ravishankar 等^[74] 利用交替优化的方法求解上述优化问题, 给出的实验结果表明 Analysis K-SVD 算法能够较好地地进行图像重建, 并且对于分段连续图像 (Piecewise-constant image) 重建更具有优势.

然而, 由于上述算法在寻找共支撑集 Λ_i 时, 需要利用预估计信号 $\hat{\mathbf{x}}_i$, 算法收敛速度较慢, 为了克服这一缺点, Zhang 等^[77] 提出了 SP-ADL (Subset pursuit for analysis dictionary learning) 算法, 该算法选择 $|\Omega \mathbf{z}_i|$ 中 $n - r$ 个最小值所对应的行索引值集合作为共支撑 $\hat{\Lambda}_i$, 然后以共支撑为依据, 对 \mathbf{z}_i 进行分类. 分类后的矩阵记为 $Z_k, \forall k \in \hat{\Lambda}_i$, 将 $Z_k Z_k^T$ 的最小特征值所对应的特征向量作为新的字典原子. SP-ADL 算法比 Analysis K-SVD 算法收敛速度更

快, 图像去噪性能更好. 结合解析模型中观测数据的共平面特点, Zhang 等提出了基于 K 平面聚类算法 K -plane-ADL (K -plane clustering algorithm for analysis dictionary learning)^[78]. 该算法不同于 SP-ADL 算法的字典更新方式, 而是采用下式更新: $\Omega_{\Lambda_i}^{(t)} = \Omega_{\Lambda_i}^{(t-1)} - \eta < \Omega_{\Lambda_i}^{(t-1)}, z_i > z_i$, 其中步长 η 随着迭代次数变化, 用来控制字典学习算法的收敛速度. Analysis K-SVD 算法每次只更新一个字典原子, 不同于该思想, Dong 等将基于综合模型的 SimCO 算法扩展到解析模型^[79], 利用简单的硬阈值算法进行稀疏编码, 在流形上采用梯度下降线性搜索方法进行字典更新. 该算法能够同时更新多个字典原子, 在较大噪声强度情况下, 去噪性能优于 Analysis K-SVD 算法, 并且字典学习速度更快.

除了上述 Analysis K-SVD 算法及其改进算法外, 目前文献公开发表的解析字典学习算法还有 Yaghoobi 等提出的约束解析算子学习 (Constrained analysis operator learning, CAOL)^[80] 算法和松弛解析算子学习 (Relaxed analysis operator learning, RAOL)^[81] 算法, Howe 等提出的几何解析算子学习 (Geometric analysis operator learning, GAOL)^[82] 算法. CAOL 算法用 l_1 范数作为 ΩX 的稀疏性度量, 用次梯度法求解相应的优化问题. 为了避免出现平凡解及某些分析原子为 0 的情况, 它们增加了 Ω 必须为均匀归一化紧标架 (Uniformly normalized tight frame, UNTF) 的约束条件, 即 Ω 每行的 l_2 范数为 1, 列与列之间相互正交. RAOL 算法的优化目标变为 Ω 在属于 UNTF 的前提下最小化 $\|\Omega X\|_1$, RAOL 的思想与 CAOL 算法类似, 只是将 Ω 属于 UNTF 的条件放宽为 Ω 与 UNTF 近似. GAOL 算法采用了与 Analysis K-SVD 完全不同的方式, 它将 Ω^T 看作是矩阵空间中斜流形 (Oblique manifold) 上的一个元素, 从而将解析字典学习问题转化为斜流形上 l_p ($0 < p < 1$) 范数优化问题. Howe 等^[82] 提出了几何共轭梯度下降法来求解斜流形上的 l_p 范数最小化问题. 上述几种算法给出的图像去噪、修复和超分辨率的实验结果显示了解析字典的巨大潜力.

3.2 稀疏变换学习算法

传统的解析稀疏变换是通过数学方法构造的, 例如小波、离散傅里叶变换等, 它们对数据的自适应能力较差. 学习稀疏变换的主要目的是通过对样本进行训练, 找到某个满足约束条件的稀疏变换 $W \in \mathbf{R}^{n \times n}$, 使得表示误差 $\|WX - A_w\|_F^2$ 最小, 其中 X 的每一列为一个训练样本, A_w 为每个样本经过变换按列组成的稀疏矩阵. 基于稀疏变换模型, Ravishankar 等提出了稀疏变换学习 (Learning

sparsifying transform, LST) 算法, 该算法对应的优化问题为

$$\begin{aligned} \min_{W, A} \{ & \|WX - A_w\|_F^2 - \lambda_w \log \det W + \\ & \beta_w \|W\|_F^2 \} \\ \text{s. t. } & \|\alpha_i\| \leq s, \quad 1 \leq i \leq N \end{aligned} \quad (32)$$

式中, λ_w 和 β_w 表示正则化参数, 代价函数的第一项保证信号在变换域误差最小, 最后两项是对稀疏变换矩阵的限制. 为了避免出现平凡解或零解, 加入了稀疏变换矩阵的范数限制, 但是这种限制并不能排除重复行的出现. 为解决这一问题, 他们将变换矩阵行列式的负对数作为惩罚项加入到代价函数中. 他们同样采用交替优化方法求解上述问题, 利用阈值法和共轭梯度法分别对稀疏系数和稀疏变换矩阵进行更新. 他们给出的基于压缩感知的核磁共振成像实验说明学习稀疏变换方法比综合字典学习方法成像效果更好^[83], 速度更快, 更具有实用价值. Ravishankar 等给出了稀疏变换学习问题的近似解^[84], 并将 LST 算法推广到过完备稀疏变换的学习, 提出了 LOST (Learning overcomplete sparsifying transform) 算法^[85], 其去噪性能优于 LST 算法. 受 Ravishankar 稀疏变换学习工作的鼓舞, Eksioğlu 等^[86] 利用稀疏变换框架对 Analysis K-SVD 算法进行了改进, 提出了变换 K-SVD (Transform K-SVD, TK-SVD) 算法. TK-SVD 算法在稀疏编码步骤中直接利用硬阈值方法寻找共支撑集及稀疏系数, 在稀疏变换更新步骤中, 保留了支撑集, 并利用奇异值分解更新变换矩阵的行, 文献 [86] 中给出的基于 TK-SVD 算法的去噪实验结果比 Analysis K-SVD 算法效果好、速度快. 上述稀疏变换学习算法都基于干净信号进行训练, 然而文献 [87] 提出直接利用噪声信号进行稀疏变换, 他们限制稀疏变换矩阵为紧标架, 利用硬阈值求解稀疏系数, 在稀疏变换学习步骤中, 利用简单的梯度下降法对稀疏变换矩阵整体进行更新. 文献 [87] 的实验结果表明, 该算法去噪性能优于 LOST 算法, 并且当噪声强度较大时优于 Analysis K-SVD 算法.

3.3 结构化解析字典学习算法

对字典进行结构限制已经在综合字典的应用中表现出明显的优势^[66-68], 近两年, 结构化解析字典的研究也受到大家的关注. 非结构化解析字典学习算法由于将图像信号转换成一维信号进行解析字典学习, 导致图像的空间性质或内在的结构没有被充分利用. 为了充分捕获图像的内在结构, 文献 [88] 中直接对图像块矩阵操作, 将 $\Omega_1 X_i \Omega_2$ (其中 X_i 表示第 i 个矩阵化的训练样本, 即第 i 个图像块的矩阵形式) 的共稀疏度限制作为约束项加入到解析字典学

习的代价函数中, 提出了解析字典学习的二维稀疏模型. 在字典更新步骤中, 利用与 Analysis K-SVD 算法类似的字典更新方式分别对 Ω_1, Ω_2 进行更新. 当 Ω_1, Ω_2 固定时, 二维稀疏模型可以转化为一维稀疏模型, 此时解析字典 $\Omega = \Omega_1 \otimes \Omega_2$, 该算法的去噪效果比 Analysis K-SVD 算法有明显提升. 本质上, 该算法利用了结构化的解析字典对图像进行表示, 从而保留了图像的结构信息. 受该思想启发, Seibert 等^[89] 将二维解析字典学习推广到多维, 他们将解析字典分解为多个较小字典的张量积, 利用对数平滑函数衡量稀疏性, 并结合 GAOL 算法学习解析字典, 仿真实验给出的三维核磁共振成像性能优于基于 K-SVD 算法的核磁共振成像. 以上两种结构化解析字典算法都是基于一般意义上的解析字典学习模型. 受 Rubinstein 等^[66] 提出的双稀疏字典启发, Ravishankar 等^[90] 研究了结构化稀疏变换学习, 并提出了双稀疏变换学习算法, 假设稀疏变换矩阵具有如下结构: $W = \Theta_w \Psi_w$, 其中 Θ_w 为稀疏矩阵, Ψ_w 为固定的稀疏变换 (如 DCT 变换). 他们利用阈值法进行稀疏编码, 利用梯度法更新稀疏变换. 实验结果表明结构化稀疏变换学习算法与无结构限制的稀疏变换学习算法^[74] 相比, 去噪效果好、速度快, 而且需要较少的训练样本. 值得指出的是只有少量文献^[84-85, 90] 说明了算法的收敛性, 解析字典的收敛性分析也是未来重要的研究方向.

4 盲字典学习算法

在求解压缩感知等反问题时, 常以信号的稀疏表示作为先验知识从部分观测值 (非完备信息) 中重构原始信号. 稀疏基或字典的选择是影响重构性能的关键因素. 目前大部分 CS 重构算法利用全局字典表示信号 (即不同的信号使用同一个字典). 全局字典对于整个训练集来说是最优的, 但对于特定的待重构信号则未必是最优表示. 能否根据信号的测量值学习与待重构信号本身相适应的字典并利用该字典进行信号重构是压缩感知领域急需解决的问题. 盲压缩感知的提出正是为了解决这一难题, 它与常规 CS 不同, 只需要知道测量值和测量矩阵, 不需要事先给定正交基或字典. BCS 的信号重构问题 (4) 在数学上是一个严重的病态问题. 除 s -稀疏条件外, 必须给字典 D 施加额外的约束条件, BCS 问题才可能有唯一解. 文献 [27] 讨论了三种情况: 1) 从多个已知的正交基中根据重构残差最小准则选择一个正交基; 2) D 为双稀疏字典; 3) D 为正交基且 D 为块对角矩阵. Silva 等^[91] 研究了另外一种情况: 4) D 为多个子字典的联合, 且每个子字典的原子相互正交, 待恢复的信号只用其中的某个子字典表示. 如果 D 满足 4 个条件之一并且测量矩阵 Φ 满足相应

的条件, 则 BCS 有唯一解. 在这几种情况下, 不但对字典 D 有严格的约束, 测量矩阵 Φ 也必须满足苛刻的条件, 如第三种情况测量矩阵 Φ 必须为多个正交基的联合, 并且矩阵 ΦD 的最小线性相关的列数应该等于测量矩阵的行数加 1. 由于盲字典学习需要对测量值进行训练, 因此必须有大量的测量值作为训练样本, 在这种情况下, 测量矩阵往往是对一个图像块进行采样, 而不能对一幅图像进行整体测量 (如 MRI (Magnetic resonance imaging)), 一般情况下, 可以选满足 RIP (Restricted isometry property) 条件的高斯随机矩阵.

由于 BCS 还处在理论发展初期阶段, 到目前为止, 只有少量文献给出了直接根据测量值训练字典的具体算法, 如 CK-SVD (Compressive K-SVD)^[92] 算法和 CDL (Compressive dictionary learning)^[93] 算法, 这两种算法都是基于文献 [27] 中的第二种情况, 假设字典 D 为双稀疏字典, 即: $D = \Psi\Theta$. CK-SVD 算法通过解决以下优化问题训练原子表示矩阵:

$$\min_{\Theta, \alpha_i} \left\{ \sum_{i=1}^N \|\mathbf{y}_i - \Phi_i \Psi \Theta \alpha_i\|_2^2 \right\} \text{ s. t. } \|\alpha_i\|_0 \leq s, \forall i \quad (33)$$

其中, $\{\mathbf{y}_i\}_{i=1}^N$ 为 CS 测量值, $\Phi_i \in \mathbf{R}^{m \times n}$ ($m < n$) 为对应于第 i 块的测量矩阵, 文献 [92] 对不同的块使用不同的测量矩阵, 从而确保能够找到刻画原始信号空间的稀疏表示模型. CK-SVD 算法采用交替迭代优化方法解决以上优化问题. 固定字典 D , 对于每一个训练样本, 用 OMP 算法求解关于 α_i 的优化问题; 固定系数 α_i , 对原子表示矩阵 Θ 的每一列进行更新, 当更新 Θ 中的第 k 列后, 结合最小二乘法利用更新后的 θ_k 更新所对应的系数 α_i^k . 该方法适用于基于压缩采样的信号重建和信号分类. CDL 算法^[93] 用 l_1 范数衡量表示系数和原子表示系数的稀疏性, 解决了以下优化问题:

$$\min_{\Theta, A} \{ \|\mathbf{Y} - \Phi \Psi \Theta A\|_F^2 + \gamma_A \|A\|_1 + \gamma_\Theta \|\Theta\|_1 \} \quad (34)$$

其中, γ_A, γ_Θ 表示正则化参数. 该算法利用交替迭代方法求解上述问题, 固定字典 $D = \Psi\Theta$, 采用 Lasso (Least absolute shrinkage and selection operator)^[36] 求解稀疏表示系数. 由于 Ψ 为已知矩阵, 字典更新过程只考虑原子表示矩阵 Θ 的更新即可, 可以将字典更新问题通过误差最小准则转化为可用 Lasso 求解的最小化问题, 从而更新字典.

盲字典学习的另一个思路是不去追求唯一解, 而是寻找可能的次优解, 从而放弃 BCS 要求的 4 个严格约束条件. 基于这种思路, Ravishankar 等^[94] 提

出了 DLMRI (Dictionary learning for compressed sensing) 算法, 该算法能够通过欠采样的 k -空间数据 (部分傅里叶系数) 同时进行字典优化与图像重建, 其对应的优化问题可表示为

$$\begin{aligned} \min_{\mathbf{x}, D, A} \{ & \sum_{i=1}^N \|R_i \mathbf{x} - D \mathbf{a}_i\|_2^2 + \gamma \|F_u \mathbf{x} - \mathbf{y}\|_2^2 \} \\ \text{s. t. } & \|\mathbf{a}_i\|_0 \leq s, \forall i \end{aligned} \quad (35)$$

式中, $\mathbf{x} \in \mathbf{C}^{nN}$ 表示原始图像信号, $\mathbf{y} \in \mathbf{C}^{mN}$ 为所对应的测量值, F_u 表示下采样傅里叶编码矩阵, 矩阵 R_i 表示从信号 \mathbf{x} 中提取出第 i 块 $R_i \mathbf{x}$ 的取算子, γ 是正则化参数, A 表示稀疏表示系数的集合, 代价函数的第二项为信号的保真项. 他们通过字典学习与图像重建两个步骤对上述优化问题进行求解, 实验表明成像效果优于基于全局字典的压缩感知算法. 上述优化算法利用简单的交替优化方法进行求解, Zhang 等^[95] 利用 Split Bregman 迭代算法^[96], 提出了利用图像测量值进行字典与图像联合优化的自适应稀疏基学习算法 (Adaptive learned sparsifying basis, ALSB). ALSB 算法根据信号的稀疏表示先验进行图像重建, 利用 Split Bregman 迭代对字典学习与图像估计两个步骤交替求解, 其中字典学习采用 K-SVD 算法, 图像估计采用简单的梯度下降法. 文中给出的实验结果表明, ALSB 算法的成像效果不仅优于全局字典的重构方法, 还优于融合多种先验知识的重构算法^[97].

上述几种盲字典学习均基于综合模型, 近些年基于解析字典学习模型的压缩感知算法引起了一些学者的关注. 他们在解析字典学习模型中加入信号保真项, 对图像和解析字典进行联合优化. Ravishankar 等提出了基于稀疏变换学习方法的压缩感知算法^[83], 其对应的优化问题表示为

$$\begin{aligned} \min_{\mathbf{x}, W, \chi, A} \{ & \|W \mathbf{x}_i - \mathbf{a}_i\|_2^2 - \lambda_W \log \det W + \\ & \beta_W \|W\|_F^2 + \tau \sum_{i=1}^N \|R_i \mathbf{x} - \mathbf{x}_i\|_2^2 + \gamma_w \|F_u \mathbf{x} - \mathbf{y}\|_2^2 \} \\ \text{s. t. } & \|\mathbf{a}_i\| \leq s_i, \quad 1 \leq i \leq N \end{aligned} \quad (36)$$

其中, $\lambda_W, \beta_W, \tau, \gamma_w$ 为正则化参数, 稀疏度 s_i 表示第 i 个样本的稀疏度, χ 表示所有估计图像块的集合.

除了上述 Ravishankar 等工作外, Wormann 等^[98] 利用 Hawe 等^[82] 提出的几何解析算子学习方法, 提出了解析盲压缩感知 (Analysis based blind compressive sensing, ABCS) 算法. 该算法在几何解析算子学习的代价函数中加入信号保真项, 将图像和字典看做流形上的一个元素, 通过几何共轭梯

度法求解相应的优化问题. ABCS 算法不仅能够同时优化字典和重构图像, 而且能够处理不同类型的测量噪声.

5 基于信息复杂度的字典学习算法

在字典学习中一般使用 l_0 范数、 l_1 范数或与其近似的函数 (如 Smooth l_0 范数) 作为表示系数的稀疏性度量. Jalali 等^[28, 99] 根据“简单”信号可以用少量比特表示的原理, 利用 Kolmogorov 复杂度来统一表示信号的“简单性”, 提出了基于压缩感知的最小复杂度追踪 (Minimum complexity pursuit, MCP) 算法框架. 二元序列的 Kolmogorov 复杂度是指输出该二元序列所使用的最短通用计算机程序长度^[100]. Jalali 等将这一信息理论中的工具推广到信号处理领域, 文献 [99] 利用 Kolmogorov 复杂度最小准则解决压缩感知的重建问题, 并从理论上分析了 Kolmogorov 复杂度能够反映信号稀疏性和结构性, 证明了完美重构复杂度已知的信号所需的样本数. 理论上, 任何基于信息复杂度的字典学习算法都可以看成求解以下优化问题:

$$\min_{D, A} \left\{ \sum_{i=1}^N \|\mathbf{x}_i - D \mathbf{a}_i\|_2^2 + \lambda_K KOL^{[\bullet]_c}(\mathbf{x}_i) \right\} \quad (37)$$

其中, λ_K 为正则化参数, $KOL^{[\bullet]_c}(\mathbf{x}_i)$ 定义为信号 \mathbf{x}_i 在分辨率 c 处的 Kolmogorov 复杂度. 遗憾的是由于有限长实序列的 Kolmogorov 复杂度是不可计算的, MCP 无法实际使用. 在文献 [29] 中, Ramirez 等提出了基于最小描述长度准则的稀疏编码和字典学习算法. 字典学习的过程可看作是最佳模型的寻找过程. 最佳模型能够有效地捕获信号的规律, 准确表示信号, 从而避免信号描述过程中的冗余性. 最小描述长度 (MDL) 基于这一认识, 通过某种准则, 试图从候选模型中找到最佳的模型对信号进行最短、最有效地描述, 这里最短指的是最少的比特数. MDL 准则最早由 Rissanen^[101] 提出, 可以看作是早期的 MDL, 经过近 30 多年的发展, 逐步形成了现代的 MDL^[102], 并被应用到了去噪等领域^[103]. Ramirez 等^[29] 首次将 MDL 框架应用到字典学习上, 提出利用最小码长作为最小描述准则进行字典训练的算法. 该算法对问题 (5) 和 (6) 分别采用交替优化方式求解. 求解该优化问题相当于交替选择最佳模型: 固定字典 $D^{(t-1)}$, 从模型集合 $M(s) = \{\mathbf{a}_i \in \mathbf{R}^K, \|\mathbf{a}_i\|_0 \leq s\}$, $s = 0, \dots, K$ 中选择能够以最小码长描述训练样本 \mathbf{x}_i 的模型, 相当于

求解下优化问题:

$$\alpha_i^{(t)}(s) = \arg \min_{\alpha_i \in M(s)} L(\mathbf{x}_i, \alpha_i) \quad \text{s.t.} \quad \|\alpha_i\|_0 \leq s, \forall i \quad (38)$$

他们用码长最小化追踪算法 (Code length minimizing pursuit algorithm, COMPA) 求解上式, 不仅可以更新表示系数, 而且可以在求解过程中自适应地确定最佳稀疏度; 系数固定时, 从模型集合 $M(K) = \{(A, D), D \in \mathbf{R}^{n \times K}, \alpha_i^{(t)} \in M(s, D)\}, 1 \leq i \leq N$ 选择最佳模型, 这里 $M(s, D)$ 表示在字典 D 下的稀疏编码模型集合, 通过求解以下优化问题, 更新字典 D :

$$\min_{(A, D) \in M(K)} L(E, A, D) \quad \text{s.t.} \quad K = 0, 1, \dots \quad (39)$$

其中, $E = X - DA$ 为表示误差. 在字典更新过程中, 通过码长最小化原则选择最佳的字典大小 K . 该算法能将信号的任何先验概率模型结合到学习过程中, 通用性强, 并且具有非参数贝叶斯字典学习算法能自适应确定字典规模、稀疏度及正则化参数的优点, 学习速度比非参数贝叶斯字典学习算法更快.

6 字典学习方法应用

字典学习是一个全新的跨学科研究领域, 它综合了图像处理、统计学、信息论、机器学习、稀疏表示和压缩感知等诸多领域的理论和方法, 它的应用非常广泛. 近些年, 国内外学者将字典学习方法应用在图像去噪、图像超分辨率、压缩成像和图像分类等领域, 并取得了出色的性能.

6.1 图像去噪

图像去噪问题作为最简单的图像反问题一直是学者们研究的重点^[104-107], 基于稀疏表示的图像去噪方法主要根据图像在字典下的稀疏性先验进行去噪. 无噪图像在字典下是稀疏的, 即图像信息蕴含在少量的系数中, 这些系数值往往比较大; 而噪声用字典表示则不是稀疏的, 表示系数较小. 较大的系数值通常在稀疏编码步骤中被保留下来, 噪声对应的系数则被去除, 因此稀疏编码能够有效去除图像噪声. 传统的基于稀疏表示的图像去噪方法往往根据图像在固定稀疏基 (字典) 下的稀疏性进行去噪^[104], 自适应能力较差, 去噪性能不高. 基于样本训练的冗余字典去噪方法具有良好的自适应性^[17], 通过对大量图像块样本进行训练获得的冗余字典能够得到比固定稀疏基更好的去噪性能. 字典训练样本可以直接利用含噪图像本身获得, 也可以从图像库中抽取大量的图像块进行训练, 前者的去噪性能一般要优于后者. 直接利用含噪图像本身学习字典和去噪需要

求解以下优化问题:

$$\min_{\alpha_i, \mathbf{x}} \left\{ \lambda_x \|\mathbf{x} - \mathbf{z}\|_2^2 + \sum_{i=1}^N \beta_i \|\alpha_i\|_0 + \sum_{i=1}^N \|D\alpha_i - R_i \mathbf{x}\|_2^2 \right\} \quad (40)$$

其中, \mathbf{x} 表示干净图像, \mathbf{z} 为噪声图像, λ_x, β_i 为正则化参数. 式 (40) 的第一项为数据保真项, 第二项表示稀疏先验, 第三项表示图像块在字典 D 表示下的误差. 文献 [17] 利用交替优化方法求解上述优化问题: 1) 固定估计图像, 利用 K-SVD 算法进行字典学习; 2) 固定字典和稀疏系数, 对 \mathbf{x} 求导并令其为零来更新图像. 利用该框架, 本文给出了基于 SGK 方法和基于 K-SVD 方法的去噪性能比较. 表 1 给出了不同噪声强度下, 基于不同字典学习方法的去噪结果的对比, 并用峰值信噪比 (Peak signal to noise ratio, PSNR) 衡量算法优劣. 从表 1 中可以看到, 综合字典中基于 K-SVD 字典学习算法的去噪性能与基于 SGK 算法的去噪性能相近, 基于 SGK 方法的去噪性能更加适用于噪声强度较小的情况. 基于解析模型的 Analysis K-SVD (AK-SVD) 算法与上述两个综合字典学习算法相比效果较差. 目前基于字典学习的去噪方法已经应用到天文图像^[106]、医学图像^[107] 等领域, 并取得了良好的去噪效果.

表 1 基于三种不同字典学习方法的去噪结果对比
Table 1 The comparison of denoised results based on three different dictionary learning methods

图像	算法	噪声方差			
		10	20	30	平均
Lena	KSVD	35.49	32.39	30.45	32.78
	SGK	35.48	32.37	30.42	32.76
	AKSVD	34.74	31.48	29.58	31.93
Barbara	KSVD	34.39	30.83	28.51	31.24
	SGK	34.39	30.82	28.49	31.23
	AKSVD	33.35	29.08	26.58	29.67
Boat	KSVD	33.63	30.38	28.41	30.81
	SGK	33.63	30.36	28.39	30.79
	AKSVD	33.09	29.54	27.63	30.09
Couple	KSVD	33.49	30.02	27.93	30.48
	SGK	33.49	29.99	27.92	30.46
	AKSVD	32.99	29.11	27.12	29.74

6.2 图像超分辨率

在实际应用中, 廉价的图像采集设备 (例如手机相机等) 采集到的图像分辨率较低, 视觉效果较差, 限制了高分辨率显示设备 (例如高分辨率液晶显示

器等)的使用. 为此人们常通过低分辨率图像重建高分辨率图像. 近年来, 基于字典学习的图像超分辨率算法引起人们的广泛关注^[108-110]. Yang 等^[109-110]提出了基于联合字典学习模型的单幅图像超分辨率重建算法, 联合字典学习问题可表示为

$$\min_{D_l, D_h, \{\alpha_i^l, \alpha_i^h\}_{i=1}^N} \sum_{i=1}^N \left\{ \frac{1}{2} \|\mathbf{x}_i^l - D_l \alpha_i^l\|_2^2 + \gamma_a \|\alpha_i^l\|_1 \right\} + \left\{ \frac{1}{2} \|\mathbf{x}_i^h - D_h \alpha_i^h\|_2^2 + \gamma_a \|\alpha_i^h\|_1 \right\} \quad \text{s. t. } \alpha_i^l = \alpha_i^h \quad (41)$$

其中, D_l, D_h 分别表示低分辨率字典和高分辨率字典 (两者合并在一起称之为联合字典), α_i^h, α_i^l 表示高-低分辨率字典所对应的系数, $\mathbf{x}_i^l, \mathbf{x}_i^h$ 分别表示低分辨率和高分辨率图像块, γ_a 为正则化参数. 该算法通过字典学习方法学习出高-低分辨率字典对, 低分辨率图像块在低分辨率字典下的表示系数直接作为高分辨率图像块在高分辨率字典的表示系数, 从而重建高分辨率图像块. 该算法能够保证获得的联合字典对高分辨率和低分辨率图像表示误差之和最小, 但并不意味重构的高分辨率表示误差最小. 为克服这个局限, Yang 等根据高分辨率图像稀疏表示误差最小准则, 利用梯度下降法优化字典对, 提出了基于耦合字典学习的超分辨率算法^[111]. Lu 等^[112]研究了上述两个字典和所对应系数的内在联系, 结合非局部相似性和流形学习理论, 提出了几何结构限制稀疏编码和两步字典训练 (Two-stage dictionary training, TSDT) 算法. 该算法学习的字典原子具有非相干性的性质, 能够有效地进行超分辨率图像重建. Wang 等^[113]假设高-低分辨率图像块对应的系数存在某种稳定的线性映射关系, 这种关系反映了高-低分辨率图像之间的内在联系. 在此基础上他们提出了能够同时学习字典对和系数映射关系的半耦合字典学习 (Semi-coupled dictionary learning, SCDL) 方法. 图 3 给出了利用耦合字典和半耦合字典进行图像超分辨的示意图, 对于联合字典和耦合字典来说, 图中 F 表示恒等映射, 即 $\alpha^h = \alpha^l$; 对于半耦合字典, F 为线性映射, 即 $\alpha^h = U\alpha^l$, 其中矩阵 U 也需要通过学习算法获得.

6.3 压缩成像

压缩成像利用部分测量值和压缩感知原理进行图像重建, 已经得到了国内外学者的关注^[114-116]. 传统的压缩成像方法利用全变差或小波稀疏先验, 由于小波原子只有水平、垂直和对角线三个方向, 无法以最优的方式表示自然图像, 图像重构性能不理想, 而使用学习的字典能够有效提高图像重构性能. 字典学习方法应用于压缩成像可以通过对大量

样本训练通用字典, 也可以根据每幅图像的测量值自适应地训练字典. 后者获得的字典与待重建图像本身相适应, 其压缩成像性能一般要优于通用字典. 如 Ravishankar 等^[94]提出的 DLMRI 算法以及 Rajwade 等^[115]和 Lingala 等^[116]提出的具有盲字典学习的高光谱成像和动态磁共振成像算法都是根据测量值本身学习自适应字典, 获得了良好的重构性能. 为了说明字典学习方法应用在压缩成像中的有效性, 图 4 给出了两种基于稀疏表示的磁共振成像算法的结果, 图 4(c) 和 (d) 分别为快速复合变量分裂算法 (Fast composite splitting algorithm, FCSA)^[117]和 DLMRI 算法^[94]的重构结果. FCSA 算法联合小波稀疏性和全变差重构图像, 而 DLMRI 算法则对字典与图像同时优化, 获得的字典能够对待重建图像本身进行最优表示. 从图 4 中可以看出, DLMRI 算法重构的图像比 FCSA 重构的图像更清晰, 保留了更多的细节信息.

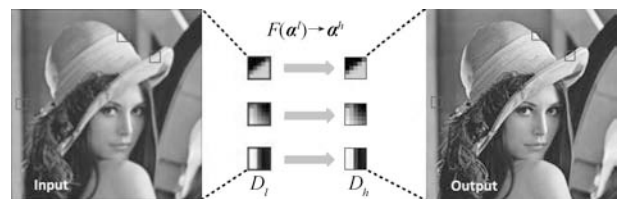


图 3 利用字典稀疏表示进行超分辨率的示意图
Fig. 3 The illustration of super-resolution based on the sparse representation

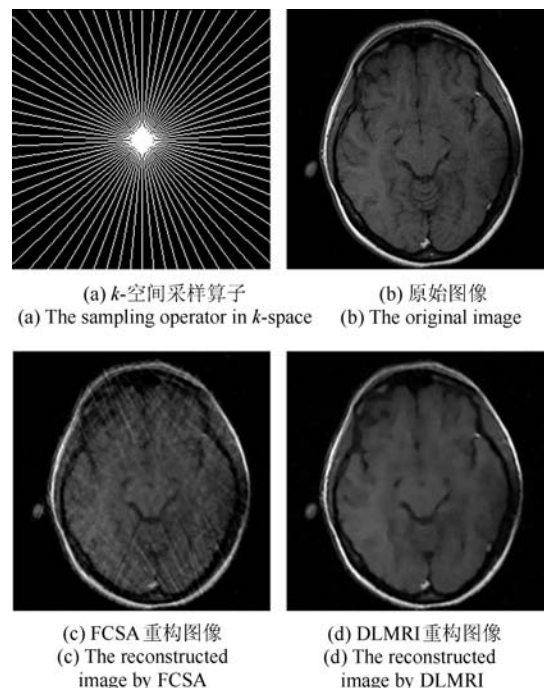


图 4 磁共振成像

Fig. 4 Magnetic resonance imaging

6.4 图像分类

基于稀疏表示的分类器 (Sparse representation-based classifier, SRC) 模型^[118] 是最早利用字典进行识别的模型, 该模型认为属于某类的测试样本主要用同类训练样本进行表示, 而不属于该类的样本表示系数应该很小, 也即测试样本在所有类别样本组成的字典上的表示系数应该是稀疏的. 基于这个原理, SRC 模型首先利用由所有测试样本组成的字典对测试样本进行稀疏编码, 然后根据重构误差最小准则判定测试样本属于哪一类.

字典学习方法应用于图像分类可以分为两种方式. 一种方式是直接使字典本身具有判别性, 另一种方式是使字典表示系数具有判别性. Yang 等在文献 [119] 中提出一种 Metaface 字典学习方法, 它对每一类样本学习一个子字典, 将各类子字典级联形成大字典, 该字典即为判别字典. 该方法基于 SRC 框架, 能用更少的原子达到较高的识别率. 为了使不同类别的字典具有非相干性, Ramirez 等^[120] 将子字典的非相干约束加入到字典学习的代价函数中, 有效降低了子字典间的相干性, 取得了较好的分类效果. 基于字典学习的图像分类的第二种方式是把稀疏系数看作新的表示特征用于分类. 这种方式只需要训练一个整体字典, 而不必明确各个类别的字典. 基于该方式, 学者们往往在字典学习时将分类误差加入到代价函数^[121-122] 中, 从而使训练好的字典 (准确说应该是字典表示系数) 具有良好的分类能力, 例如 Mairal 等^[121] 结合逻辑回归模型提出的有监督字典学习方法、Zhang 等^[122] 结合线性分类器的性能和字典表示能力提出的判别性 KSVD (Discriminative K-SVD, D-KSVD) 字典学习方法、Yang 等^[123] 结合 Fisher 准则学习结构化字典提出的 Fisher 判别字典学习方法等. 然而以上方法都是针对样本的原始空间, 训练的样本空间也可以推广到更高维的特征空间, Nguyen 等^[124] 利用 Mercer 核函数, 在高维特征空间学习了具有良好分类效果的字典, 为字典学习应用在图像分类提供了一个新的思路.

7 结论与展望

本文从综合字典学习、解析字典学习、盲字典学习以及基于信息复杂度的字典学习四大方面所涉及的重要算法及相关模型进行了详细阐述, 并介绍了字典学习在相关领域的应用实例. 综合字典学习模型发展比较成熟, 本文对基于该模型的不同算法进行了归类. Sparsenet 算法和 MOD 字典学习算法都是对整个字典进行更新, 而 K-SVD 算法对字典原子逐个更新, 并且联合优化稀疏系数和字典原子, 加

快了收敛速度. 以上三种算法及其对应的改进算法都是批处理字典学习算法, 即对所有训练样本进行计算. 为了能够在样本动态到达的情况下也能进行字典学习, ODL 算法能够每次只处理一个样本, 利用一个样本对字典进行更新, 节省了存储空间. 约束字典和结构字典的提出加快了算法的收敛速度, 对于特定的应用场合具有重要意义. 非参数贝叶斯字典学习模型的主要优点是不用手动或交叉验证设定模型参数和正则化参数, 在实际应用中更具有优势. 解析模型通过字典训练试图找到能够表示信号的正交补空间, 表示能力更强. 学习稀疏变换模型作为一种广义的解析模型, 比传统 Cospase 解析模型表示能力更强, 而且基于稀疏变换模型的稀疏变换学习速度更快, 在压缩成像和去噪应用中更具优势. 结构化解析字典缩小了最优字典的搜索范围, 比无结构限制的解析字典训练速度快、所需样本少, 更适用于对结构数据进行训练. 盲字典学习模型主要针对压缩感知问题提出的字典学习模型, 它是通过对字典进行结构限制, 利用测量值学习与待重构图像本身相适应的字典, 能有效提高图像重构性能. 信息复杂度模型利用信息论的信息复杂度衡量稀疏性, 与编码理论紧密相关, 它具有自动更新字典规模、稀疏度和正则化参数的优点. 信息复杂度模型从信息论的角度衡量信号的“简单性”, 信息复杂度可以看作是一种广义的稀疏性度量.

上述国内外研究成果表明世界各国对稀疏表示和字典学习研究的高度重视, 并且取得了很多原创成果, 但在该领域仍有不少问题没有解决.

1) 大规模字典学习的加速

随着大数据时代的来临, 数据处理速度越来越受到人们的关注, 而字典学习由于需要大量的数据计算不能满足人类发展的需求. 在保证训练字典性能的情况下, 大规模字典学习加速问题是目前亟待解决的一个难题. 研究字典结构和数据结构是解决加速问题的一种途径, 例如相对于批量处理的字典学习算法^[20, 41], 在线字典学习^[22] 和结构字典学习算法^[23] 速度有明显提升. 但是, 当面临大规模字典学习问题时, 上述加速算法仍然不能满足人们的需求. 利用图形处理器 (Graphics processing unit, GPU) 强大的并行处理功能提高数据处理速度至关重要. 但是 GPU 处理数据必须要求数据处理任务是并行的, 如何使字典学习方法具有良好的并行性是解决加速问题的一大突破点. 目前, 学者们已经将并行处理方式应用于 KSVD 字典学习方法^[44], 字典学习速度有了很大提升. 并行方式应用于更多的字典学习方法从而加速大规模字典学习是以后的一个重要研究方向.

2) 任务驱动字典的学习

传统的字典学习大都是数据驱动字典 (Data-driven dictionary) 学习方法, 利用数据的结构特性进行字典学习, 大都属于无监督学习的范畴. 虽然无监督字典学习方法被应用在了广泛的领域, 例如图像分类、压缩成像等, 但是, 一些学者已经证明了根据具体任务调整的字典能够获得更好的效果, 例如针对图像分类任务提出的有监督字典学习算法^[120]、针对压缩感知任务提出的测量矩阵与字典联合优化算法^[125-126]. 任务驱动字典 (Task-driven dictionary) 学习在特定任务处理中显示出了巨大的生命力, 如何根据具体任务学习字典是未来研究的热点之一.

3) “简单性”的度量

稀疏性是“简单性”的一种, 并且使用最为广泛, 大多数字典学习方法度量信号的稀疏性通过 l_0 范数或者 l_p ($0 < p < 1$) 范数^[127-129]. 近年来, 一些学者用矩阵的低秩性度量信号的“简单性”^[130], 并取得了良好的性能. 低秩性可以看成矩阵奇异值构成向量的稀疏性, 利用低秩性和稀疏性统一度量也越来越受到学者们的关注, 文献 [131] 研究了稀疏性和低秩性结合模型求解反问题的基本理论, 文献 [132] 综述了压缩感知到低秩矩阵恢复的理论和应用, 稀疏性和低秩性的统一度量必将受到越来越多的学者重视. 度量“简单性”除了这两个性质外, 还有低熵等, 将这些“简单性”的度量方式统一并应用在图像反问题求解中, 通过具体与抽象相结合的方式描述“简单性”, 很可能会出现一个新的理论分支^[133], 具有很高的研究价值.

4) 字典学习收敛性的分析

字典学习问题本身是一个非凸优化问题, 虽然有大量算法求解该问题, 但是很少涉及算法的收敛性分析. 字典学习算法的收敛性分析至关重要, 不仅可以作为评价算法优劣的标准, 而且通过研究收敛性, 很有可能使最终解突破局部最优的限制, 达到全局最优. 目前, 字典学习算法的收敛性研究还很少, 只有少量公开发表的文献^[39, 134-135] 从理论上证明了所提出的字典学习算法的收敛性, 例如文献 [135] 对稀疏系数的大小进行限制, 利用邻近方法解决了基于 l_0 范数的字典学习问题, 并给出了全局收敛性的证明. 字典学习算法的收敛性分析是今后字典学习理论层面的一个重要研究方向.

5) 解析字典的学习

相对于综合模型, 解析模型从另一角度考虑问题, 具有更丰富的表示能力, 已经引起了学者们的关注, 并取得了一些阶段性成果^[74-87]. 但是, 现阶段基于解析模型的字典学习方法仍属于理论发展初期, 如何精确快速地寻找共支撑是需要解决的基本问题, 目前, Zhang 等^[77-78] 已经对 Rubinstein 等^[75] 寻

找共支撑的方式进行了改进, 但仍有很大的提升空间. 解析字典学习的应用成果还比较少, 目前国内外研究主要应用在图像去噪^[77-78] 和压缩成像^[83, 98] 领域, 将这一方法应用于其他领域是今后的一个发展方向. 由于图像中存在平滑、边缘、纹理、轮廓等各层面的结构特征, 根据图像信号本身的结构特点研究结构化解析字典学习方法^[88-90], 也是未来解析字典学习的一个发展方向. 作为广义的解析字典学习, 稀疏变换学习算法比一般意义的解析字典学习算法具有速度快、性能高的优点, 以后必将得到更多学者的关注.

6) 盲字典的学习

盲字典学习通过对测量值进行训练, 联合优化字典与表示系数, 从而重构图像, 它以良好的自适应性在压缩感知领域有很好的应用前景. 盲字典学习还处于理论发展的初级阶段, 目前只有 Gleichman 等和 Silva 等提出的 4 种情况^[27, 91] 能保证信号唯一性重建. 在实际应用中, 测量矩阵很多情况下是由物理条件决定的, 很难满足这些苛刻条件, 信号重建的唯一性条件仍需要突破. 通过研究稀疏表示的结构特性, 放松 BCS 限制的苛刻要求, 得到新的唯一性重建条件, 是盲字典学习理论层面需要解决的问题. 另外, 后三种情况是通过对应的字典学习算法求解 BCS 问题, 因此测量值矩阵的列数必须足够大. 如果在一幅图像进行整体测量 (如 MRI), 则只能得到一个测量向量 (即 \mathbf{Y} 是向量而不是矩阵), BCS 无法应用在这种场合. 因此 BCS 适用范围还很窄, 它只适合于信号分段压缩感知或图像分块压缩感知系统, 并且对测量矩阵及字典有严格的限制. 如何突破这些限制, 将盲字典学习方法应用于更多的领域也是未来的研究方向.

References

- Hubel D H, Wiesel T N. Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology*, 1959, **148**(3): 574-591
- Willshaw D J, Buneman O P, Longuet-Higgins H C. Non-holographic associative memory. *Nature*, 1969, **222**(5197): 960-962
- Barlow H B. Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1972, **1**(4): 371-394
- Oja E. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 1982, **15**(3): 267-273
- Pham T T, Defigueiredo R J P. Maximum likelihood estimation of a class of non-Gaussian densities with application to

- L_p deconvolution. *IEEE Transaction on Acoustics, Speech, and Signal Process*, 1989, **37**(1): 73–82
- 6 Jutten C, Herault J. Blind separation of sources, Part I: an adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 1991, **24**(1): 1–10
- 7 Mallat S, Zhang Z. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 1993, **41**(12): 3397–3415
- 8 Chen S S, Donoho D L, Saunders M A. Atomic Decomposition by Basis Pursuit. Technical Report, Stanford University, Britain, 1995.
- 9 Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996, **381**(6583): 607–609
- 10 Olshausen B A, Field D J. Natural image statistics and efficient coding. *Network Computation in Neural Systems*, 1996, **7**(2): 333–339
- 11 Olshausen B A, Field D J. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Research*, 1997, **37**(23): 3311–3325
- 12 Donoho D L. Compressed sensing. *IEEE Transactions on Information Theory*, 2006, **52**(4): 1289–1306
- 13 Candes E J, Tao T. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Transactions on Information Theory*, 2006, **52**(12): 5406–5425
- 14 Liu Fang, Wu Jiao, Yang Shu-Yuan, Jiao Li-Cheng. Research advances on structured compressive sensing. *Acta Automatica Sinica*, 2013, **39**(12): 1980–1995
(刘芳, 吴娇, 杨淑媛, 焦李成. 结构化压缩感知进展. 自动化学报, 2013, **39**(12): 1980–1995)
- 15 Mallat S. Geometrical grouplets. *Applied and Computational Harmonic Analysis*, 2009, **26**(2): 161–180
- 16 Yaghoobi M, Daudet L, Davies M E. Parametric dictionary design for sparse coding. *IEEE Transactions on Signal Processing*, 2009, **57**(12): 4800–4810
- 17 Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Signal Processing*, 2006, **15**(12): 3736–3745
- 18 Liu J J, Ma X H. An improved image inpainting algorithm based on multi-scale dictionary learning in wavelet domain. In: Proceedings of the 2013 International Conference on Signal Processing, Communication and Computing. Kunming, China: IEEE, 2013. 1–5
- 19 Liu X M, Zhai D M, Zhao D B, Gao W. Image super-resolution via hierarchical and collaborative sparse representation. In: Proceedings of the 2013 Data Compression Conference. Snowbird, USA: IEEE, 2013. 93–102
- 20 Aharon M, Elad M, Bruckstein A M. The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 2006, **54**(11): 4311–4322
- 21 Yaghoobi M, Blumensath T, Davies M E. Dictionary learning for sparse approximations with the majorization method. *IEEE Transactions on Signal Processing*, 2009, **57**(6): 2178–2191
- 22 Mairal J, Bach F, Ponce J, Sapiro G. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 2010, **11**(1): 19–60
- 23 Zelnik-Manor L, Rosenblum K, Eldar Y C. Dictionary optimization for block-sparse representations. *IEEE Transactions on Signal Processing*, 2012, **60**(5): 2386–2395
- 24 Elad M, Milanfar P, Rubinstein R. Analysis versus synthesis in signal priors. *Inverse Problems*, 2007, **23**(3): 947–968
- 25 Rubinstein R, Bruckstein A M, Elad M. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 2010, **98**(6): 1045–1057
- 26 Nam S, Davies M E, Elad M, Gribonval R. The cosparsity analysis model and algorithms. *Applied and Computational Harmonic Analysis*, 2013, **34**(1): 30–56
- 27 Gleichman S, Eldar Y C. Blind compressed sensing. *IEEE Transactions on Information Theory*, 2011, **57**(10): 6958–6975
- 28 Jalali S, Maleki A. Minimum complexity pursuit. In: Proceedings of the 49th Annual Allerton Conference on Communication, Control, and Computing. Monticello, IL: IEEE, 2011. 1764–1770
- 29 Ramirez I, Sapiro G. An MDL framework for sparse coding and dictionary learning. *IEEE Transactions on Signal Processing*, 2012, **60**(6): 2913–2927
- 30 Donoho D L, Tsaig Y, Drori I, Starck J L. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 2012, **58**(2): 1094–1121
- 31 Shi Guang-Ming, Liu Dan-Hua, Gao Da-Hua, Liu Zhe, Lin Jie, Wang Liang-Jun. Advances in theory and application of compressed sensing. *Acta Electronica Sinica*, 2009, **37**(5): 1070–1081
(石光明, 刘丹华, 高大化, 刘哲, 林杰, 王良君. 压缩感知理论及其研究进展. 电子学报, 2009, **37**(5): 1070–1081)

- 32 Tropp J A, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 2007, **53**(12): 4655–4666
- 33 Dai W, Milenkovic O. Subspace pursuit for compressive sensing signal. *IEEE Transaction on Information Theory*, 2009, **55**(5): 2230–2249
- 34 Ambat S K, Chatterjee S, Hari K V S. Fusion of algorithms for compressed sensing. *IEEE Transactions on Signal Processing*, 2013, **61**(14): 3699–3704
- 35 Chen S S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 1998, **20**(1): 33–61
- 36 Tibshirani R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B*, 1996, **58**(1): 267–288
- 37 Lewicki M S, Olshausen B A. Probabilistic framework for the adaptation and comparison of image codes. *Journal of the Optical Society of America a Optics Image Science and Vision*, 1999, **16**(7): 1587–1601
- 38 Kreutz-Delgado K, Murray J F, Rao B D, Engan K, Lee T W, Sejnowski T J. Dictionary learning algorithms for sparse representation. *Neural Computation*, 2003, **15**(2): 349–396
- 39 Mailhé B, Plumbley M D. Dictionary learning with large step gradient descent for sparse representations. In: Proceedings of the 10th International Conference on Latent Variable Analysis and Signal Separation. Berlin, Heidelberg: Springer, 2012. 231–238
- 40 Lee H, Battle A, Raina R, Ng A Y. Efficient sparse coding algorithms. In: Proceedings of the 20th Annual Conference on Neural Information Processing Systems. Columbia, Canada: IEEE, 2006. 801–808
- 41 Engan K, Aase S O, Husoy J H. Method of optimal directions for frame design. In: Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Phoenix AZ: IEEE, 1999. 2443–2446
- 42 Smith L N, Elad M. Improving dictionary learning: multiple dictionary updates and coefficient reuse. *IEEE Signal Processing Letters*, 2013, **20**(1): 79–82
- 43 Rubinstein R, Zibulevsky M, Elad M. Efficient Implementation of the K-SVD Algorithm Using Batch Orthogonal Matching Pursuit, Technical Report, Technion University, Israel, 2008.
- 44 Sadeghi M, Babaie-Zadeh M, Jutten C. Learning overcomplete dictionaries based on parallel atom-updating. In: Proceedings of the 2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP). Southampton, UK: IEEE, 2013. 1–5
- 45 Dai W, Xu T, Wang W. Simultaneous codeword optimization (SimCO) for dictionary update and learning. *IEEE Transactions on Signal Processing*, 2012, **60**(12): 6340–6353
- 46 Rusu C, Dumitrescu B. Stagewise K-SVD to design efficient dictionaries for sparse representations. *IEEE Signal Processing Letters*, 2012, **19**(10): 631–634
- 47 Lu C W, Shi J P, Jia J Y. Scale adaptive dictionary learning. *IEEE Transactions on Image Processing*, 2014, **23**(2): 837–847
- 48 Sahoo S K, Makur A. Dictionary training for sparse representation as generalization of K-means clustering. *IEEE Signal Processing Letters*, 2013, **20**(6): 587–590
- 49 Sadeghi M, Babaie-Zadeh M, Jutten C. Dictionary learning for sparse representation: a novel approach. *IEEE Signal Processing Letters*, 2013, **20**(12): 1195–1198
- 50 Rakotomamonjy A. Applying alternating direction method of multipliers for constrained dictionary learning. *Neurocomputing*, 2013, **106**: 126–136
- 51 Rakotomamonjy A. Direct optimization of the dictionary learning problem. *IEEE Transactions on Signal Processing*, 2013, **61**(22): 5495–5506
- 52 Sigg C D, Dikk T, Buhmann J M. Learning dictionaries with bounded self-coherence. *IEEE Signal Processing Letters*, 2012, **19**(12): 861–864
- 53 Mailhe B, Barchiesi D, Plumbley M D. INK-SVD: learning incoherent dictionaries for sparse representations. In: Proceedings of the 2012 International Conference on Acoustics, Speech and Signal Processing. Kyoto, Japan: IEEE, 2012. 3573–3576
- 54 Barchiesi D, Plumbley M D. Learning incoherent dictionaries for sparse approximation using iterative projections and rotations. *IEEE Transactions on Signal Processing*, 2013, **61**(8): 2055–2065
- 55 Skretting K, Engan K. Recursive least squares dictionary learning algorithm. *IEEE Transactions on Signal Processing*, 2010, **58**(4): 2121–2130
- 56 Labusch K, Barth E, Martinetz T. Robust and fast learning of sparse codes with stochastic gradient descent. *IEEE Transactions on Selected Topics in Signal Processing*, 2011, **5**(5): 1048–1060

- 57 Sendur L, Selesnick I W. Bivariate shrinkage functions for wavelet based denoising exploiting interscale dependency. *IEEE Transactions on Signal Processing*, 2002, **50**(11): 2744–2756
- 58 Lesage S, Gribonval R, Bimbot F, Benaroya L. Learning unions of orthonormal bases with thresholded singular value decomposition. In: Proceedings of the 2005 International Conference on Acoustics, Speech and Signal Processing. Philadelphia, PA: IEEE, 2005. 293–296
- 59 Vidal R, Ma Y, Sastry S. Generalized principal component analysis (GPCA). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(12): 1945–1959
- 60 Bengio S, Pereira F, Singer Y, Strelow D. Group sparse coding. In: Proceedings of the 23rd Annual Conference on Neural Information Processing Systems. Vancouver, British Columbia, Canada: IEEE, 2009. 82–89
- 61 Szabo Z, Poczos B, Lorincz A. Online group-structured dictionary learning. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI: IEEE, 2011. 2865–2872
- 62 Jenatton R, Mairal J, Obozinski G, Bach F. Proximal methods for hierarchical sparse coding. *Journal of Machine Learning Research*, 2011, **12**(7): 2297–2334
- 63 Mairal J, Sapiro G, Elad M. Learning multiscale sparse representations for image and video restoration. *Multiscale Modeling and Simulation*, 2008, **7**(1): 214–241
- 64 Ophir B, Lustig M, Elad M. Multi-scale dictionary learning using wavelets. *IEEE Journal of Selected Topics in Signal Processing*. 2011, **5**(5): 1014–1024
- 65 Thiagarajan J J, Ramamurthy K N, Spanias A. Multilevel dictionary learning for sparse representation of images. In: Proceedings of the 2011 IEEE Digital Signal Processing Workshop and Signal Processing Education Workshop. Sedona, AZ: IEEE, 2011. 271–276
- 66 Rubinstein R, Zibulevsky M, Elad M. Double sparsity: learning sparse dictionaries for sparse signal approximation. *IEEE Transactions on Signal Processing*, 2010, **58**(3): 1553–1564
- 67 Yaghoobi M, Davies M E. Compressible dictionary learning for fast sparse approximations. In: Proceedings of the 15th IEEE/SP Workshop on Statistical Signal Processing. Cardiff: IEEE, 2009. 662–665
- 68 Hawe S, Seibert M, Kleinsteuber M. Separable dictionary learning. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE, 2013. 438–445
- 69 Jost P, Vandergheynst P, Lesage S, Gribonval R. MoTIF: an efficient algorithm for learning translation invariant dictionaries. In: Proceedings of the 2006 IEEE International Conference on Acoustics, Speech and Signal Processing. Toulouse, France: IEEE, 2006. 5
- 70 Aharon M, Elad M. Sparse and redundant modeling of image content using an image-signature-dictionary. *SIAM Journal on Imaging Sciences*, 2008, **1**(3): 228–247
- 71 Rusu C, Dumitrescu B, Tsiftaris S A. Explicit shift-invariant dictionary learning. *IEEE Signal Processing Letters*, 2014, **21**(1): 6–9
- 72 Pope G, Aubel C, Studer C. Learning phase-invariant dictionaries. In: Proceedings of the 2013 International Conference on Acoustics, Speech and Signal Processing. Vancouver, Canada: IEEE, 2013. 5979–5983
- 73 Zhou M, Yang H, Paisley J, Ren L. Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images. *IEEE Transactions on Image Processing*, 2012, **21**(1): 130–144
- 74 Ravishanker S, Bresler Y. Learning sparsifying transform. *IEEE Transactions on Signal Processing*, 2013, **61**(5): 1072–1086
- 75 Rubinstein R, Peleg T, Elad M. Analysis K-SVD: a dictionary-learning algorithm for the analysis sparse model. *IEEE Transactions on Signal Processing*, 2013, **61**(3): 661–677
- 76 Chen Y J, Ranftl R, Pock T. Insights into analysis operator learning: from patch-based sparse models to higher order MRFs. *IEEE Transactions on Image Processing*, 2014, **23**(3): 1060–1072
- 77 Zhang Y, Wang H L, Yu T L. Subset pursuit for analysis dictionary learning. In: Proceedings of the 21th European Signal Processing Conference. Marrakech, Morocco, 2013. 1–5
- 78 Zhang Y, Wang H L, Wang W W, Sanei S. K-plane clustering algorithm for analysis dictionary learning. In: Proceedings of the 2013 IEEE International Workshop on Machine Learning for Signal Processing. Southampton, Britain: IEEE, 2013. 1–4
- 79 Dong J, Wang W W, Dai W. Analysis SimCO: a new algorithm for analysis dictionary learning. In: Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing. Florence, Italy: IEEE, 2014. 7193–7197
- 80 Yaghoobi M, Nam S, Gribonval R, Davies M E. Constrained overcomplete analysis operator learning for cosparse signal

- modelling. *IEEE Transactions on Signal Processing*, 2013, **61**(9): 2141–2355
- 81 Yaghoobi M, Davies M E. Relaxed analysis operator learning. In: Proceedings of the 2012 NIPS, Workshop on Analysis Operator Learning vs. Dictionary Learning: Fraternal Twins in Sparse Modeling. Lake Tahoe, USA: IEEE, 2012.
- 82 Hawe S, Kleinstueber M, Diepold K. Analysis operator learning and its application to image reconstruction. *IEEE Transactions on Image Processing*, 2013, **22**(6): 2138–2150
- 83 Ravishanker S, Bresler Y. Sparsifying transform learning for compressed sensing MRI. In: Proceedings of the 10th International Symposium on Biomedical Imaging. San Francisco, CA: IEEE, 2013. 17–20
- 84 Ravishanker S, Bresler Y. Closed-form solutions within sparsifying transform learning. In: Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, Canada: IEEE, 2013. 5378–5382
- 85 Ravishanker S, Bresler Y. Learning sparsifying transforms for image processing. In: Proceedings of the 2012 19th IEEE International Conference on Image Processing. Orlando, USA: IEEE, 2012. 681–684
- 86 Eksioğlu E M, Bayir O. K-SVD meets transform learning: transform K-SVD. *IEEE Signal Processing Letters*, 2014, **21**(3): 347–351
- 87 Zhang Y, Yu T L, Wang W W. An analysis dictionary learning algorithm under a noisy data model with orthogonality constraint. [Online], available: <http://www.hindawi.com/journals/tswj/2014/852978/>, June 26, 2014
- 88 Qi N, Shi Y H, Sun X Y, Wang J D, Ding W P. Two dimensional analysis sparse model. In: Proceedings of the 20th International Conference on Imaging Processing. Melbourne, Australia: IEEE, 2013. 310–314
- 89 Seibert M, Wörmann J, Gribonval R, Kleinstueber M. Separable cospase analysis operator learning. In: Proceedings of the 22nd European Signal Processing Conference (EU-SIPCO). Lisbonne, Portuga, 2014.
- 90 Ravishanker S, Bresler Y. Learning doubly sparsifying transforms for images. *IEEE Transactions on Image Processing*, 2013, **22**(12): 4598–4612
- 91 Silva J, Chen M, Eldar Y C, Sapiro G, Carin L. Blind compressed sensing over a structured union of subspaces. [Online], available: <http://webee.technion.ac.il/people/YoninaEldar/BlindC.pdf>, 2011
- 92 Anaraki P F, Hughes S M. Compressive K-SVD. In: Proceedings of the 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing. Vancouver, Canada: IEEE, 2013. 5469–5473
- 93 Aghagolzadeh M, Radha H. Compressive dictionary learning for image recovery. In: Proceedings of the 19th IEEE International Conference on Image Processing. Orlando, USA: IEEE, 2012. 661–664
- 94 Ravishanker S, Bresler Y. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Transactions on Medical Imaging*, 2011, **30**(5): 1028–1041
- 95 Zhang J, Zhao C, Zhao D B, Gao W. Image compressive sensing recovery using adaptively learned sparsifying basis via L_0 minimization. *Signal Processing*, 2014, **103**: 114–126
- 96 Goldstein T, Osher S. The split Bregman algorithm for L_1 regularized problems. *SIAM Journal on Imaging Sciences*, 2009, **2**(2): 323–343
- 97 Chen C, Tramel E W, Fowler J E. Compressed-sensing recovery of images and video using multi-hypothesis predictions. In: Proceedings of the 45th Asilomar Conference on Signals, Systems, and Computers. Pacific Grove, USA: IEEE, 2011. 1193–1198
- 98 Wormann J, Hawe S, Kleinstueber M. Analysis based blind compressive sensing. *IEEE Signal Processing Letters*, 2013, **20**(5): 491–494
- 99 Jalali S, Maleki A, Baraniuk R G. Minimum complexity pursuit for universal compressed sensing. *IEEE Transactions on Information Theory*, 2014, **60**(4): 2253–2268
- 100 Kolmogorov A N. Logical basis for information theory and probability theory. *IEEE Transactions on Information Theory*, 1968, **14**(5): 662–664
- 101 Rissanen J. Modeling by shortest data description. *Automatica*, 1978, **14**(5): 465–471
- 102 Grunwald P D. *The Minimum Description Length Principle*. Cambridge: UK Press, 2007.
- 103 Roos T, Myllymaki P, Rissanen J. MDL denoising revisited. *IEEE Transactions on Signal Processing*, 2009, **57**(9): 3347–3360
- 104 Chang S G, Yu B, Vetterli M. Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 2000, **9**(9): 1532–1546
- 105 Knaus C, Zwicker M. Dual-domain image denoising. In: Proceedings of the 2013 20th IEEE International Conference on Image Processing. Melbourne, Australia: IEEE, 2013. 440–444

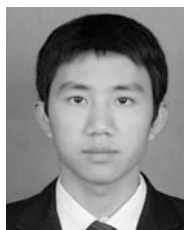
- 106 Beckouche S, Starck J L, Fadili J. Astronomical image denoising using dictionary learning. *Astronomy & Astrophysics*, 2013, **556**(6): 14, DOI: 10.1051/0004-6361/201220752
- 107 Li S T, Fang L Y, Yin H T. An efficient dictionary learning algorithm and its application to 3-D medical image denoising. *IEEE Transactions on Biomedical Engineering*, 2012, **59**(2): 417–427
- 108 Lian Qiu-Sheng, Zhang Jun-Qin, Chen Shu-Zhen. Single image super-resolution algorithm based on two-stage and multi-frequency-band dictionaries. *Acta Automatica Sinica*, 2013, **39**(8): 1310–1320
(练秋生, 张钧芹, 陈书贞. 基于两级字典与分频带字典的图像超分辨率算法. 自动化学报, 2013, **39**(3): 1310–1320)
- 109 Yang J C, Wright J, Huang T, Ma Y. Image super-resolution as sparse representation of raw image patches. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE, 2008. 1–8
- 110 Yang J C, Wright J, Huang T S, Ma Y. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 2010, **19**(11): 2861–2873
- 111 Yang J C, Wang W Z, Lin Z, Cohen S, Huang T. Coupled dictionary training for image super-resolution. *IEEE Transaction on Image Processing*, 2012, **21**(8): 3467–3478
- 112 Lu X Q, Yuan Y, Yan P K. Alternatively constrained dictionary learning for image superresolution. *IEEE Transactions on Cybernetics*, 2014, **44**(3): 366–377
- 113 Wang S L, Zhang L, Liang Y, Pan Q. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012. 2216–2223
- 114 Lian Qiu-Sheng, Chen Shu-Zhen. Image reconstruction for compressed sensing based on the combined sparse image representation. *Acta Automatica Sinica*, 2010, **36**(3): 385–391
(练秋生, 陈书贞. 基于混合基稀疏图像表示的压缩传感图像重构. 自动化学报, 2010, **36**(3): 385–391)
- 115 Rajwade A, Kittle D, Tsai T H, Brady D, Carin L. Codes hyperspectral imaging and blind compressive sensing. *SIAM Journal on Imaging Sciences*, 2013, **6**(2): 782–812
- 116 Lingala S G, Jacob M. A blind compressive sensing framework for accelerated dynamic MRI. In: Proceedings of the 9th IEEE International Symposium on Biomedical Imaging (ISBI). Barcelona, Spain: IEEE, 2012. 1060–1063
- 117 Huang J Z, Zhang S T, Metaxas D. Efficient MR image reconstruction for compressed MR imaging. *Medical Image Analysis*, 2011, **15**(5): 670–679
- 118 Wright J, Yang A, Ganesh A, Sastry S, Ma Y. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence*, 2009, **31**(2): 210–227
- 119 Yang M, Zhang L, Yang J, Zhang D. Metaface learning for sparse representation based face recognition. In: Proceedings of the 17th IEEE International Conference on Image Processing. Hong Kong, China: IEEE, 2010. 1601–1604
- 120 Ramirez I, Sprechmann P, Sapiro G. Classification and clustering via dictionary learning with structured incoherence and shared features. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 3501–3508
- 121 Mairal J, Bach F, Ponce J. Task-driven dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(4): 791–804
- 122 Zhang Q, Li B Q. Discriminative K-SVD for dictionary learning in face recognition. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 2691–2698
- 123 Yang M, Zhan D, Feng X C, Zhang D. Fisher discrimination dictionary learning for sparse representation. In: Proceedings of the 2011 IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011. 543–550
- 124 Nguyen H V, Patel V M, Nasrabadi N M, Chellapa R. Kernel dictionary learning. In: Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing Kyoto, Japan: IEEE, 2012. 2021–2024
- 125 Duarte-Carvejalino J M, Sapiro G. Learning to sense sparse signals: simultaneous sensing matrix and sparsifying dictionary optimization. *IEEE Transactions on Image Processing*, 2009, **18**(7): 1395–1408
- 126 Chen W, Rodrigues M R D. Dictionary learning with optimized projection design for compressive sensing applications. *IEEE Signal Processing Letters*, 2013, **20**(10): 992–995
- 127 Zhang Hai, Wang Yao, Chang Xiang-Yu, Xu Zong-Ben. L1/2 regularization. *Science China: Information Sciences*, 2010, **40**(3): 412–422
(张海, 王尧, 常象宇, 徐宗本. L1/2 正则化. 中国科学: 信息科学, 2010, **40**(3): 412–422)
- 128 Xu Z B, Chang X Y, Xu F M, Zhang H. L1/2 regularization: a thresholding representation theory and a fast solver. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, **23**(7): 1013–1027

- 129 Zuo W M, Meng D Y, Zhang L, Feng X C, Zhang D. A generalized iterated shrinkage algorithm for non-convex sparse coding. In: Proceedings of the 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013. 217–244
- 130 Candes E J, Plan Y. Matrix completion with noise. *Proceedings of the IEEE*, 2010, **98**(6): 925–936
- 131 Candès E J, Recht B. Simple bounds for recovering low-complexity models. *Mathematical Programming*, 2013, **141**(1–2): 577–589
- 132 Peng Yi-Gang, Suo Jin-Li, Dai Qiong-Hai, Xu Wen-Li. From compressed sensing to low-rank matrix recovery: theory and applications. *Acta Automatica Sinica*, 2013, **39**(7): 981–994
(彭义刚, 索津莉, 戴琼海, 徐文立. 从压缩传感到低秩矩阵恢复: 理论与应用. 自动化学报, 2013, **39**(7): 981–994)
- 133 Elad M. Sparse and redundant representation modeling—what next? *IEEE Signal Processing Letters*, 2012, **19**(12): 922–928
- 134 Mailhe B, Plumbley M D. Fixed points of dictionary learning algorithms for sparse representations. [Online], available: <http://hal.inria.fr/docs/00/80/75/45/PDF/locOpt.pdf>, April 3, 2013
- 135 Bao C L, Ji H, Quan Y H, Shen Z W. L_0 norm based dictionary learning by proximal methods with global convergence. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, USA: IEEE, 2013. 3858–3865



练秋生 燕山大学信息科学与工程学院教授. 主要研究方向为图像处理, 稀疏表示, 压缩感知及多尺度几何分析. 本文通信作者. E-mail: lianqs@ysu.edu.cn

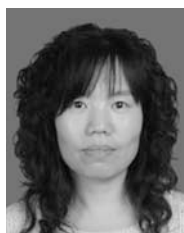
(LIAN Qiu-Sheng Professor at the School of Information Science and Engineering, Yanshan University. His research interest covers image processing, sparse representation, compressed sensing, and multi-scale geometrical analysis. Corresponding author of this paper.)



石保顺 燕山大学信息科学与工程学院博士研究生. 主要研究方向为图像处理, 盲压缩感知, 字典学习.

E-mail: shibaoshun1989@163.com

(SHI Bao-Shun Ph. D. candidate at the School of Information Science and Engineering, Yanshan University. His research interest covers image processing, blind compressed sensing, and dictionary learning.)



陈书贞 燕山大学信息科学与工程学院副教授. 主要研究方向为图像处理, 压缩感知及生物识别.

E-mail: chen_sz818@163.com

(CHEN Shu-Zhen Associate professor at the School of Information Science and Engineering, Yanshan University. Her research interest covers image processing, compressed sensing, and biometrics recognition.)